

84254

From: Chan, Christina
Sent: Tuesday, January 14, 2003 3:04 PM
T : Yu, Misook; STIC-Biotech/ChemLib
Subject: RE: rush search for 09/648,310

Importance: High

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

CRF

-----Original Message-----

From: Yu, Misook
Sent: Tuesday, January 14, 2003 11:43 AM
To: Chan, Christina
Subject: rush search for 09/648,310

Please approve rush searh. The case is due this bi-week.

Please search SEQ ID NO:1 and 2.

Examiner Misook YU, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

Point of Contact:
Barb O'Bryen
Technical Information Specialist
STIC CM1 6A05 308-4291

Searcher: 1003
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 1-17-03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

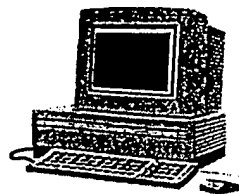
TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

BioTech-Chem Library

Search Results

Feedback Form (Optional)



Scientific & Technical Information Center

The search results generated for your recent request are attached. If you have any questions or comments (compliments or complaints) about the scope or the results of the search, please contact *the BioTech-Chem searcher* who conducted the search *or contact*:

Mary Hale, Supervisor, 308-4258
CM-1 Room 1E01

Voluntary Results Feedback Form

➤ *I am an examiner in Workgroup:* (Example: 1610)

➤ *Relevant prior art found, search results used as follows:*

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ *Relevant prior art not found:*

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Search results were not useful in determining patentability or understanding the invention.

Other Comments:

Drop off completed forms at the Circulation Desk CM-1, or send to Mary Hale, CM1-1E01 or mary.hale@uspto.gov

Yu, Misook

To: Chan, Christina
Subject: rush search for 09/648,310

Please approve rush search. The case is due this bi-week.

Please search SEQ ID NO:1 and 2.

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)


```

QY 543 TATACCTTGGAGCAAAATACATGTAATAAATAAACAACCTGTTATTTTTTTTTC 602
Db 253 TTTTATATCTTGAAGAAATCTATGATGCTATATAAAATAATCCTATTTTTC 194
QY 603 TTAAGAAAGTAAATCGGAGAGCTAGGCAATATAATGTTTTTCAGAGTCCGAAAAAGCTTT 662
Db 193 TCAGGAATCTGGTAGGAATTCAGCAATAGAGATTTTTCGGGGCAAGATGGGAATG 134
QY 663 TGTTTCTTAAACCATCTTCTAGTCT--CTGCCACATGAGCACTCCGTCAAAAGTGAGAAG 720
Db 133 TTTGTTCAATAAATAATAGACATTTTCTATAGATATTGACATTTCTGGAAGCAACAAG 74
QY 721 CGAACTAAAGACCAACTCGCGTGGAAATATATGTTTATGTAATAAAAAA 773
Db 73 CAACCTGAAGACCAACTCCTATGAGAAATATATGATGTTTATGTAATAAGA 21

```

RESULT 4

```

US-09-925-299-318
; Sequence 318, Application US/09925299
; Patent No. US20020053627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL02
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 318
; LENGTH: 388
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (3)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (14)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (143)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (144)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (146)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (150)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (154)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (159)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (160)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (163)

```

```

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (168)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (171)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (174)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (196)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (198)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (223)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (226)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (227)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (234)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (240)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (242)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (255)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (280)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (314)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (359)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (366)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (368)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (377)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-299-318

```

```

Query Match 15.7%; Score 122.4; DB 10; Length 388;
Best Local Similarity 65.8%; Pred. NO. 1.4e-23;
Matches 156; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 109 GGTTCAGACCCACGCGGAGCAGCTCTTCAGTGAAGAGGAGCAATCGGAGGTCAGC 168
Db 140 GGTNNACCATNGGGGNNAGNTCTCNGGNGNAGGAAATCGTCAAGAGGNANT 199
QY 169 AATGAAGCTGGAGCATGAGGTTAACTCTCTGTCGAGGAAATTCATCTCGGTTCCAA 228
Db 200 AATGAATGTGGATCAGGAGTANCCNNTTGTGNAGGAANTNCATCGTTGGGNTTAA 259
QY 229 AATGCCGATGGGAAACTGAGTGAAGTTTGGGTCCTCTTCCAGAGCAGAGATGTC 288
Db 260 AATGCTGATGGAAGTTAANTGTGAATTTGGGTCCTCTTTCGTGATGATANTGAGC 319

```



```

Query Match      8.1%; Score 63.2; DB 10; Length 289;
Best Local Similarity 59.3%; Pred. No. 1e-07;
Matches 121; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

QY 167 GCAATGAACCTGGAGCATGAGGTTAAACCTCCTGGTGGAGGAAATTCATCGTCTGGGTTCC 226
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 GCATGAACCTGGAGGATGAGGTCGGGAGGCTCAANGAGGAGATCCATAGCTCGGCCAG 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 227 AAAAAATGCCGATGGGAACATGAGTGTGAAGCTTTGGGGTCTCTTCCAAAGACGACAGATGT 286
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 CAGCAGCCCGATNGCTCTTACAAGGTCAAGCTTTGGCGTTCTCTTCAACGACGATCGGTGT 195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 287 GCCAAATCTCTTTGAAGCGTTGGTGGGAACCTCTGAAGCCCGAAAACGAAAGAAATTTGT 346
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 GCAATATC-TTGAAGCACATAGTTGGCACTTGAAGGCGCCAGAGAGTAGATTTGACT 254
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 347 ACGTACGACGAGAGCTGCTTTTG 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 ACGAAGCGGANTGTTCTGCAAGTG 278
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-10-098-841-105/c
; Sequence 105, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungling
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679alel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2
; CURRENT APPLICATION NUMBER: US/10/098,841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: PL_FL_genes Version 1.0
; SEQ ID NO 105
; LENGTH: 6222
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (707)..(3379)
US-10-098-841-105

```

	Query Match	4.9%	Score 38.6;	DB 9;	Length 6222;
	Best Local Similarity	54.6%;	Pred. No. 1.9;		
	Matches	77;	Conservative	0;	Mismatches 64; Indels 0; Gaps 0;
Qy	457	TTAAGTTAAAGGCAACACATCAAGTTCCCTATGATGATATTTTATAGACCTTTGTAAACAAA	516		
Db	3751	TTAAAAAACAACAAACAACATATATTTATATTTTATATATATATATATATATATAA	3692		

```

Qy 517 GGGGACTGTTGAGAAGTCCTGTTTTTATCCTTGGAGCAAAACATTTACAATGTAAAAAT 576
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3691 AGAGAAATGGGAAGACAAAATTAAGTGAGCCTATTGTATTAAAGTGAAGGATTAAAGAAAA 3632
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 577 AAACAAAACCTGTTATTTTTT 597
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3631 ATACAAAACCAAGTATTGCTT 3611
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-08-927-939-30
; Sequence 30, Application US/08927939
; Patent No. US20010006640A1
; GENERAL INFORMATION:
; APPLICANT: Grainger, David J.
; APPLICANT: Tatalick, Lauren Marie
; TITLE OF INVENTION: Compounds and methods to inhibit or
; TITLE OF INVENTION: augment an inflammatory response.
; FILE REFERENCE: 295.022US1
; CURRENT APPLICATION NUMBER: US/08/927,939
; CURRENT FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 30
; LENGTH: 4259
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2070)...(2130)
; NAME/KEY: CDS
; LOCATION: (2669)...(2795)
; NAME/KEY: CDS
; LOCATION: (2990)...(3079)
; NAME/KEY: CDS
; LOCATION: (3491)...(3506)
US-08-927-939-30

```

	Query Match	4.9%;	Score 38.4;	DB 8;	Length 4259;
	Best Local Similarity	48.6%;	Pred. No.1.8;		
	Matches 105;	Conservative	0;	Mismatches 111;	Indels 0;
	Gaps				
Qy	443	GTAACCTGGGAATTAATTAAGTTAAAGGACAAACATGAAGTCCCTTATGTATTTTATATAGAC	502		
Db	4003	GTAARTACTATTTANTGAACGACTGTACAAAGTAGAATTCCTAGATGATATTTTTTGTATG	4062		
Qy	503	CTTTGTTAAACAAAAGGGGACTTGTGAGAGTCCCTGTTTTTATACCTTGGAGCAAAACAT	562		
Db	4063	CTTTGCAATTGTATATGATGGAAGAACTTGTGTCATCAAGTATGTATCAATGGTAGTTAAAGT	4132		
Qy	563	TACAATGTAAAAATAACAAACACCTCTTATTTTTTTTTTTCTTAAAGAGGTATCGGGAGA	622		
Db	4123	TTATTTTTAAACCGTCCAATACCTTTTGTATATGTACATTCAAAGACANTGACTG	4182		
Qy	623	CGTAGGCAATAAATGTTTTTCAGAGGTGCGAAAAAG	658		
Db	4183	TATTGAAAGTAGTAAGAGACCCCAAAATGTAAATAAG	4218		

RESULT 13
US-09-938-842A-3235
; Sequence 3235, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRP1300-3
; CURRENT APPLICATION NUMBER: US/09/938, 842A
; CURRENT FILING DATE: 2001-08-24

;; PRIOR APPLICATION NUMBER: US 60/227,866
;; PRIOR FILING DATE: 2000-08-24
;; PRIOR APPLICATION NUMBER: US 60/264,647
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: US 60/300,111
;; PRIOR FILING DATE: 2001-06-22
;; NUMBER OF SEQ ID NOS: 5379
;; SEQ ID NO 3235
;; LENGTH: 2000
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3235

Query Match 4.9%; Score 38; DB 9; Length 2000;
Best Local Similarity 47.8%; Pred. No. 1.6;
Matches 110; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
QY 448 CTGGAATAATTAAGCTTAAGGACAAACATGAAGTTCCCTATGTATTTTATAGACCTTTG 507
DB 25 CTTTAAATCTATTGTAATTAACCTTTTAAATCCCTTGTAAATCTTTATAAACCCCTA 84
QY 508 TAAACAAAGGGGACTTGTGAGAGTCCTGTTTTTATACCTTGGAGCAAAACATTACAA 567
DB 85 TAAATCCCTGTGAATCTTTTAAACTCTTTTAAATCCCTTGTAAATCTTTATAAACCCCTA 144
QY 568 TGTAAATAAATAACAAACCTGTTATTTTTTTTCTTAAAGAGTAAATCGGGAGACGTAG 627
DB 145 CAAATAAAATAAATTTGGCGGAAAAATATGTTGAAAAACAAATTTTGGCGGAAATTTAT 204
QY 628 GCAATAAAATGTTTTCAGAGTGTGCGAAAGCTTTGTTTCTTAAACCA 677
DB 205 TTGGAATATTTTGGCGAAAAATATTAATTTTCATTTTGTAAACCTA 254

RESULT 14
US-10-046-935-342/c
;; Sequence 342, Application US/10046935
;; Patent No. US20020156011A1
;; GENERAL INFORMATION:
;; APPLICANT: Jiang, Yuqiu
;; APPLICANT: Harlocker, Susan L.
;; APPLICANT: Secrist, Heather
;; APPLICANT: Wang, Aijun
;; APPLICANT: Stolk, John A.

;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
;; FILE REFERENCE: 210121.527C1
;; CURRENT APPLICATION NUMBER: US/10/046,935
;; NUMBER OF SEQ ID NOS: 2239
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 342
;; LENGTH: 338
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-046-935-342

Query Match 4.8%; Score 37.6; DB 9; Length 338;
Best Local Similarity 53.4%; Pred. No. 0.85;
Matches 79; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 489 GTATTTTATAGACCTTTGTAACAAAGGGGACTTCTTGAGAGTCTCTGTTTATACC 548
DB 325 GTTATATATAATAGTGTAGAAAAATTTTATTATTAAACAGTCTTTTGGTCCC 266
QY 549 TTGGAGCAAAACATTACAATGTAATAATAACAAACCTGTTATTTTTTTCTTAAGA 608
DB 265 ATCAACACATATCTTCTAGCAATAATAACAAACATCTTTTTTTTTTTTGGCT 206
QY 609 AGGTAATCGGGAGACGTAGGCAATAAAA 636
DB 205 AGGCTTCTCCAAATCGTGGGAATAAACA 178

RESULT 15
US-09-878-178-342/c
;; Sequence 342, Application US/09878178
;; Patent No. US20020177552A1
;; GENERAL INFORMATION:
;; APPLICANT: Jiang, Yuqiu
;; APPLICANT: Harlocker, Susan L.
;; APPLICANT: Secrist, Heather
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
;; FILE REFERENCE: 210121.527
;; CURRENT APPLICATION NUMBER: US/09/878,178
;; NUMBER OF SEQ ID NOS: 2237
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 342
;; LENGTH: 338
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-09-878-178-342

Query Match 4.8%; Score 37.6; DB 9; Length 338;
Best Local Similarity 53.4%; Pred. No. 0.85;
Matches 79; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 489 GTATTTTATAGACCTTTGTAACAAAGGGGACTTCTTGAGAGTCTCTGTTTATACC 548
DB 325 GTTATATATAATAGTGTAGAAAAATTTTATTATTAAACAGTCTTTTGGTCCC 266
QY 549 TTGGAGCAAAACATTACAATGTAATAATAACAAACCTGTTATTTTTTTCTTAAGA 608
DB 265 ATCAACACATATCTTCTAGCAATAATAACAAACATCTTTTTTTTTTTTGGCT 206
QY 609 AGGTAATCGGGAGACGTAGGCAATAAAA 636
DB 205 AGGCTTCTCCAAATCGTGGGAATAAACA 178

Search completed: January 17, 2003, 02:04:09
Job time : 112 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2003, 21:41:43 ; Search time 71 Seconds
(without alignments)
3369.125 Million cell updates/sec

Title: US-09-648-310-1
Perfect score: 780
Sequence: 1 ggcacagctctctctgtcc.....gtaataaaaaaaatcatgt 780

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	157	20.1	177	2	US-09-032-684-20
2	44	5.6	7218	1	US-08-232-463-14
3	40	5.1	611	4	US-09-385-982-376
4	38	4.9	720	1	US-08-117-083-23
5	38	4.9	1500	1	US-08-117-083-67
6	37	4.7	1818	4	US-09-357-206A-6
7	37	4.7	5253	4	US-09-357-206A-16
8	37	4.7	5483	4	US-09-357-206A-17
9	37	4.7	5586	4	US-09-357-206A-19
10	37	4.7	5816	4	US-09-357-206A-21
11	37	4.7	6095	4	US-09-357-206A-18
12	37	4.7	6325	4	US-09-357-206A-20
13	37	4.7	6428	4	US-09-357-206A-22
14	37	4.7	7400	1	US-08-261-563A-1
15	37	4.7	7400	5	PCT-US95-07754A-1
16	37	4.7	10881	4	US-09-357-206A-9
17	37	4.7	12286	4	US-09-357-206A-1
18	36.8	4.7	3588	1	US-08-197-792-32
19	36.8	4.7	3588	1	US-08-459-850-32
20	36.8	4.7	3588	1	US-08-459-214-32
21	35.4	4.5	907	1	US-08-664-596B-23
22	35.4	4.5	907	2	US-08-739-775-1
23	35	4.5	4880	4	US-09-402-929-5
24	35	4.5	168575	4	US-09-426-290-1
25	34.4	4.4	4291	2	US-08-417-210A-81
26	34.4	4.4	5852	4	US-07-867-106-2
27	34.2	4.4	2981	1	US-08-257-073-2

28 34.2 4.4 2981 2 US-08-184-009-119 Sequence 119, App
29 34.2 4.4 2981 2 US-08-458-356-119 Sequence 119, App
30 34.2 4.4 2981 4 US-08-460-736-119 Sequence 119, App
31 34.2 4.4 3107 4 US-08-213-419B-1 Sequence 1, Appl
32 34.2 4.4 5506 4 US-09-004-838-93 Sequence 93, Appl
33 34.2 4.4 6124 4 US-08-213-419B-3 Sequence 3, Appl
34 34.2 4.4 7641 1 US-07-924-028A-5 Sequence 5, Appl
35 34.2 4.4 10607 1 US-08-078-090-3 Sequence 3, Appl
36 34.2 4.4 1713 4 US-09-134-001C-135 Sequence 135, App
37 33.6 4.3 7218 1 US-08-232-463-14 Sequence 14, Appl
38 33.4 4.3 1511 1 US-07-991-867B-8 Sequence 8, Appl
39 33.4 4.3 1511 1 US-08-107-755A-8 Sequence 8, Appl
40 33.4 4.3 1511 2 US-08-544-332-8 Sequence 8, Appl
41 33.4 4.3 1511 4 US-09-370-861A-8 Sequence 8, Appl
42 33.2 4.3 606 4 US-09-328-111-132 Sequence 132, App
43 33 4.2 1208 4 US-09-461-474-11 Sequence 11, Appl
44 33 4.2 2196 1 US-08-090-523-29 Sequence 29, Appl
45 33 4.2 2196 1 US-08-334-639-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-032-684-20
; Sequence 20, Application US/09032684
; Patent No. 5882874
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: RECIPROCAL SUBTRACTION DIFFERENTIAL
; TITLE OF INVENTION: DISPLAY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,684
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 55551/JPW/AMG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-032-684-20

Query Match 20.1%; Score 157; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 9e-37;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 GTAGCAATAAATGTTTCAGAGTGCGAAAAGCTTTTGTCTTAAACCACTCTTA 683
|||||

Db 1 GTAGCAATAAATGTTTCAGAGTGCGAAAAGCTTTTGTCTTAAACCACTCTTA 60
|||||

QY 684 GTCTCTGCCACACTTGACACTCCGTCAAAGTGAGAGCGCACTAAAGACCACTCGGTG 743
|||||

Db 61 GTCTCTCCACACTTGACACTCGCTAAAGTGAGAGCACTAAAGACCACTGCGGTG 120
QY 744 GAAATATATGTTATGTAATAAAAAAATCATGT 780
Db 121 GAAATATATGTTATGTAATAAAAAAATCATGT 157

RESULT 2

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFELINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match 5.6%; Score 44; DB 1; Length 7218;
Best Local Similarity 14.1%; Pred. No. 0.0066;
Matches 80; Conservative 214; Mismatches 274; Indels 0; Gaps 0;

QY 78 TTCCTCCCTAGGCGGGAAGCTGAGTGCAGGTTTCAGACCCAGCGGCGAGCAGCTCTT 137
Db 1441 TTGCTACRNR 1382
QY 138 CAGTGAGAGGAGGAGCAATCGGAGGTGACCAATGACGTGGAGCATGAGGTTAACCTC 197
Db 1381 RNR 1322
QY 198 TGGTGAGGAAATTCATCGTCTGGTCCAAAATCCGATGGGAACTGAGTGTGAGT 257
Db 1321 RNR 1262

QY 258 TTGGGGTCTCTTCCAGACGACAGATGTGCCAATCTCTTTGAAGGTTGGTGGGAATC 317
Db 1261 RNR 1202
QY 318 TCAAGACCCCAACGACGAGATTTGTCAGTACGACGAGAGCTGCTTTTCCAGGTG 377
Db 1201 RNR 1142
QY 378 TTCATGATGATGTGACATTTGCTCTCAAGATTAAATGCTGCTTTCAGATCTGGGGT 437
Db 1141 RNR 1082
QY 438 ATCTGTTAACTGGAATAATTAAGTTAAAGGACAAACATGAAGTCTCTTATGTTT 497
Db 1081 RNR 1022
QY 498 TAGACCTTTGTAACAAAAGGGGACTTTGTGAGAACTCTCTGTTTATACCTTTGGAGCA 557
Db 1021 TGAGCGTATGCAAAACGAGGAAATAGTTATAGTACGCGCACTCGATGGGACATTTCA 962
QY 558 AACATTACAATGTAATAATAACAAAACCTGTTATTTTTTTTCTTAAGAGGTAATCG 617
Db 961 ACGTAAACCGTTTAAATAATTTTGAATCTTATTCATTATCTGAAATGTTGTAATACT 902
QY 618 GGAGACGTAGGCAATAAATAATTTTTCAG 645
Db 901 AACTGCTGCTGATGCAAAATGCTTTAAG 874

RESULT 3

US-09-385-982-376
; Sequence 376, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CDDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 376
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(611)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-376

Query Match 5.1%; Score 40; DB 4; Length 611;
Best Local Similarity 62.6%; Pred. No. 0.029;
Matches 107; Conservative 0; Mismatches 54; Indels 10; Gaps 3;

QY 436 GTATCTGTAACACTGGAATAATTAAGTTAAAGGACAAACAT---GAGTTCCTTATGAT 492
Db 325 GTTTTCNGGTAACNGGAAT-ATAANGNGAAGAACAAACNTTTGGAACATCTAATGAT 383
QY 493 TTTTATAGACCTTTTAAACAAAAGGAGCTGT-----TCAGAGTCTCTCTTTTATA 546
Db 384 TTTTATAGAACTTTGNAACCAAGAGGATTCATGTTTANAAGTCTGGCCTTTTATA 443
QY 547 CCTTGGAGCAAAACATTAACAATGTAATAAATAACAAACCTGTTATTTT 597
Db 444 TCTTGGAGAAATATCTGTTNTGGAGCTTAAATAATATCCCATTTTCT 494

```

RESULT 4
US-08-117-083-23
; Sequence 23, Application US/08117083
; Patent No. 5719054
; GENERAL INFORMATION:
; APPLICANT: Bournnell, Michael E.
; APPLICANT: Inglis, Stephen C.
; APPLICANT: Munro, Alan J.
; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human
; TITLE OF INVENTION: Papilloma Virus Proteins
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Walter H. Dreger
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,083
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-58783
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-761-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 720 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-117-083-23

```

Query Match	4.9%;	Score 38;	DB 1;	Length 720;
Best Local Similarity	49.5%;	Pred. No. 0.12;		
Matches 98;	Conservative 0;	Mismatches 100;	Indels 0;	Gaps 0;
QY 351	ACGCAGGAGAGCTGCTTTTGGCAAGGTGTCATGATGATGTTGACATTTGCTGCGCAAG	410		
Db 296	ATGCTGTTGTGCTGTTTGCAGATTGGCCAGATCATGGTATATGGATACATAATGGTAT	355		
QY 411	ATTAATGTGGTTTGCAGATCTGGGGGTATCTGGTAAACTTGGAAATAATTAAGTTAAAGGAC	470		
Db 356	CAAGAAGTATATTCTCCAGATGAATGGGTGCACATATAAAATTTTAAATTAATGTAATAGAG	415		
QY 471	AAACATGAAGTTCCTTATGTAATTTTATAGACCTTTGTAACAAAGGGGACTTGTTCAG	530		
Db 416	AACAAATAAATGTTGTAATATCATATAGACAATACTAAACAATTAATTAGTAACCTGT	475		
QY 531	AAGTCCTGTTTTTTTATACC	548		
Db 476	ATCTCTTTTTTTTAACTAAC	493		

RESULT 5
US-08-117-083-67
; Sequence 67, Application US/08117083
; Patent No. 5719054
; GENERAL INFORMATION:
; APPLICANT: Bournsnel, Michael E.

APPLICANT: Inglis, Stephen C.
 APPLICANT: Munro, Alan J.
 TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human
 TITLE OF INVENTION: Papilloma Virus Proteins
 NUMBER OF SEQUENCES: 70
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Walter H. Dreger
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/117,083
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Dreger, Walter H.
 REGISTRATION NUMBER: 24,190
 REFERENCE/DOCKET NUMBER: A-58783
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-781-1989
 TELEFAX: 415-398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 67:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1500 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-117-083-67

Query Match 4.9%; Score 38; DB 1; Length 1500;
Best Local Similarity 49.5%;
Matches 98; Conservative 0; Mismatches 100; Indels
Pred. NO. 0.18;

Qy	351	ACG	CAGGAGAC	TGCTTTT	TGCAAGGT	GTTCATGA	GAATGTTG	GACATTTG	TATCTG	CTC	CAAG	410
Db	743	ATG	TGTTG	TGCTG	TGTTTG	CAGAGT	TGCCAGAA	TATCATG	TATG	TACTA	TATG	802
Qy	411	ATT	AAATG	TGGTTT	CCAGAT	CTGGGGGT	TATCTG	TGTAAC	TGTAAT	TAAATG	TTAAAGG	470
Db	803	CAAG	AAGTAT	TCTCC	AGATGA	ATGGGTG	TCACAT	TATAAA	ATTTAA	TATG	TAAATAG	862
Qy	471	AAAC	ATGAAG	TTCC	TATG	TATATTTT	TATAGAC	CTTTGT	TAACAA	CAAAAGG	GGGAC	530
Db	863	AACA	AAATAA	AAAG	TTG	TATAT	TATCAT	TATAG	CAACA	TAACTA	AACA	922
Qy	531	AAG	TCTG	TGTTTTT	TATACC	548						
Db	923	ATC	TCTTTTTT	TAACTAAC	940							

RESULT 6

US-09-357-206A-6/c
: Sequence 6, Application US/09357206A
: Patent No. 6372962
: GENERAL INFORMATION:
: APPLICANT: Dinesh-Kumar, S.
: APPLICANT: Baker, Barbara
: TITLE OF INVENTION: Pathogen Resistance
: FILE REFERENCE: 042250/191805 (5830-5)
: CURRENT APPLICATION NUMBER: US/09/357,206A
: CURRENT FILING DATE: 1999-07-20
: PRIOR APPLICATION NUMBER: US 60/093,494
: PRIOR FILING DATE: 1998-07-20
: NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 6

; LENGTH: 1818

; TYPE: DNA

; ORGANISM: Nicotiana glauca

US-09-357-206A-6

Query Match 4.7%; Score 37; DB 4; Length 1818;

Best Local Similarity 50.7%; Pred. No. 0.38;

Matches 115; Conservative 0; Mismatches 110; Indels 2; Gaps 1;

QY 390 TTGACATTTGTTGCTGCAAGATTAAATGTTGCGAGATCTGGGGTATCTGTAACACT 449

Db 1490 TTGAATACAGAGGCCAACACATTTATTTGTTGTAATAGAAATTTATTTTATATA 1431

QY 450 GGAATAATTA--AGTTAAAGGACAAACATGAAGTCTCTATGTTATTTTATAGACCTTTG 507

Db 1430 TGGAATAATTTACTTTTAAGAACTGAATTAAGGTAATTTATTTATATATATATA 1371

QY 508 TAAACAAAGGGGACTGTTGAGAAGTCTGTTTATACCTTGGAGCAAAACATTACAA 567

Db 1370 AAATTACAAAGGCTCTTAATGTCGTTCTTTTACCCTTTTAAATAAATTTTACAC 1311

QY 568 TGTAAATAATAACAAACCTGTTATTTTCTTTTCTTAAGAAGTAA 614

Db 1310 TAGAGAAATATATTTTAAGTTAAATTTCTTAATATTTAGGATTTA 1264

RESULT 7

US-09-357-206A-16/c

; Sequence 16, Application US/09357206A

; Patent No. 6372962

; GENERAL INFORMATION:

; APPLICANT: Dinesh-Kumar, S.

; APPLICANT: Baker, Barbara

; TITLE OF INVENTION: Pathogen Resistance in Plants using cDNA-N/Intron Constructs

; FILE REFERENCE: 042250/191805 (5830-5)

; CURRENT APPLICATION NUMBER: US/09/357,206A

; CURRENT FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: US 60/093,494

; PRIOR FILING DATE: 1998-07-20

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 16

; LENGTH: 5253

; TYPE: DNA

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: cDNA-N/intron construct: E1-E2-E3-I3-E4-E5

US-09-357-206A-16

Query Match

Best Local Similarity 4.7%; Score 37; DB 4; Length 5253;

Matches 115; Conservative 0; Mismatches 110; Indels 2; Gaps 1;

QY 390 TTGACATTTGTTGCTGCAAGATTAAATGTTGCGAGATCTGGGGTATCTGTAACACT 449

Db 3338 TTGAATACAGAGGCCAACACATTTATTTGTTGTAATAGAAATTTATTTTATATA 3279

QY 450 GGAATAATTA--AGTTAAAGGACAAACATGAAGTCTCTATGTTATTTTATAGACCTTTG 507

Db 3278 TGGAATAATTTACTTTTAAGAACTGAATTAAGGTTAAATTTATTTATATATATA 3219

QY 508 TAAACAAAGGGGACTGTTGAGAAGTCTTTTATACCTTGGAGCAAAACATTACAA 567

Db 3218 AAATTACAAAGGCTCTTAATGTCGTTCTTTTACCCTTTTAAATAAATTTTACAC 3159

QY 568 TGTAAATAATAACAAACCTGTTATTTTCTTTTCTTAAGAAGTAA 614

Db 3158 TAGAGAAATATATTTTAAGTTAAATTTCTTAATATTTAGGATTTA 3112

RESULT 8

US-09-357-206A-17/c

; Sequence 17, Application US/09357206A

; Patent No. 6372962

; GENERAL INFORMATION:

; APPLICANT: Dinesh-Kumar, S.

; APPLICANT: Baker, Barbara

; TITLE OF INVENTION: Pathogen Resistance in Plants using cDNA-N/Intron Constructs

; FILE REFERENCE: 042250/191805 (5830-5)

; CURRENT APPLICATION NUMBER: US/09/357,206A

; CURRENT FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: US 60/093,494

; PRIOR FILING DATE: 1998-07-20

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 17

; LENGTH: 5483

; TYPE: DNA

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: cDNA-N/intron construct: E1-E2-E3-I3-E4-E5

US-09-357-206A-17

Query Match 4.7%; Score 37; DB 4; Length 5483;

Best Local Similarity 50.7%; Pred. No. 0.65;

Matches 115; Conservative 0; Mismatches 110; Indels 2; Gaps 1;

QY 390 TTGACATTTGTTGCTGCAAGATTAAATGTTGCGAGATCTGGGGTATCTGTAACACT 449

Db 3568 TTGAATACAGAGGCCAACACATTTATTTGTTGTAATAGAAATTTATTTTATATA 3509

QY 450 GGAATAATTA--AGTTAAAGGACAAACATGAAGTCTCTTATGTTATTTTATAGACCTTTG 507

Db 3508 TGGAATAATTTACTTTTAAGAACTGAATTAAGGTAATTTATTTATATATATA 3449

QY 508 TAAACAAAGGGGACTGTTGAGAAGTCTCTGTTTATACCTTGGAGCAAAACATTACAA 567

Db 3448 AAATTACAAAGGCTCTTAATGTCGTTCTTTTACCCTTTTAAATAAATTTTACAC 3389

QY 568 TGTAAATAATAACAAACCTGTTATTTTCTTTTCTTAAGAAGTAA 614

Db 3388 TAGAGAAATATATTTTAAGTTAAATTTCTAATATTTAGGATTTA 3342

RESULT 9

US-09-357-206A-19/c

; Sequence 19, Application US/09357206A

; Patent No. 6372962

; GENERAL INFORMATION:

; APPLICANT: Dinesh-Kumar, S.

; APPLICANT: Baker, Barbara

; TITLE OF INVENTION: Pathogen Resistance in Plants using cDNA-N/Intron Constructs

; FILE REFERENCE: 042250/191805 (5830-5)

; CURRENT APPLICATION NUMBER: US/09/357,206A

; CURRENT FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: US 60/093,494

; PRIOR FILING DATE: 1998-07-20

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 19

; LENGTH: 5586

; TYPE: DNA

; ORGANISM: artificial

; FEATURE:

; OTHER INFORMATION: cDNA-N/intron construct: E1-E2-E3-I3-E4-I4-E5

US-09-357-206A-19

Query Match 4.7%; Score 37; DB 4; Length 5586;

Best Local Similarity 50.7%; Pred. No. 0.65;

Matches 115; Conservative 0; Mismatches 110; Indels 2; Gaps 1;

QY 390 TTGACATTTGTTGCTGCAAGATTAAATGTTGCGAGATCTGGGGTATCTGTAACACT 449

Db 3338 TTGAATACAGAGGCCAACACATTTATTTGTTGTAATAGAAATTTATTTTATATA 3279

[illegible]

```

RESULT 10
US-09-357-206A-21/c
; Sequence 21, Application US/09357206A
; Patent No. 6372962
; GENERAL INFORMATION:
; APPLICANT: Dinesh-Kumar, S.
; APPLICANT: Baker, Barbara
; TITLE OF INVENTION: Pathogen Resistance in Plants using cDNA-N/Intron Constructs
; FILE REFERENCE: 042250/191805 (5830-5)
; CURRENT APPLICATION NUMBER: US/09/357,206A
; CURRENT FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: US 60/093,494
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 5816
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: cDNA-N/Intron construct: E1-I1-E2-E3-I3-E4-I4-E5
US-09-357-206A-21

```

Query Match	4.7%;	Score 37;	DB 4;	Length 5816;
Best Local Similarity	50.7%;	Pred. No. 0.67;		
Matches 115;	Conservative	0;	Mismatches 110;	Indels 2; Gaps 1;
QY	390	TTGCATCTGTATTCTGCCAAGATTAAATCTGGTTTGCAGATCTGGGGGTATCTCGTAAACT	449	
Db	3568	TTGCATTTACGAGGCCCAACACATTTATTTTGTGTGAATATAGAAATTTATTTTATATATA	3509	
QY	450	GGATAAATA - AGTTAAAGGACAAACATGAAGTTCCTTATGTATTTTATAGACCTTTG	507	
Db	3508	TGGAAAAATTACTTTCTTGAAGAACTGAATAAAGCTAAATTTATTTATTTACTATATTA	3449	
QY	508	TAAACAAAGGGGACTTGTTCAGAAGTCCTCTTTTATATACCTTCGGACAAACATTCACAA	567	
Db	3448	AAATTACAAGGCTCTTAATGTCTCTTCTTTTACCTTTTAAAAATAAATTTTACAC	3389	
QY	568	TGTAATAATAACAAAAACCTGTATTATTTTTTTTTTCTTAAAGAGGTAA	614	
Db	3388	TAGACGAAATATATTTTAACTTAATTTCTTCAATATTTAGGATTTTA	3342	

```

RESULT 11
US-09-357-206A-18/c
; Sequence 18, Application US/09357206A
; Patent No. 6372962
; GENERAL INFORMATION:
; APPLICANT: Dinesh-Kumar, S.
; APPLICANT: Baker, Barbara
; TITLE OF INVENTION: Pathogen Resistance in Plants using CDNA-N/Intron Constructs
; FILE REFERENCE: 042250/191805 (5830-5)
; CURRENT APPLICATION NUMBER: US/09/357,206A
; CURRENT FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: US 60/093,494
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0

```

```

; SEQ ID NO 18
; LENGTH: 6095
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: cdna-N/intron construct: E1-E2-12-E3-13-E4-E5
us-09-357-206A-18

Query Match      4.7%; Score 37; DB 4; Length 6095;
Best Local Similarity 50.7%; Pred. No. 0.68;
Matches 115; Conservative 0; Mismatches 110; Indels 2; Gaps

Qy 390 TTGACATTGTATTGCTGCCAAGATTAAATGTGGTTTCAGATCTGGGGGATCTGGTAAACT 449
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4180 TTGAATTACGAGGGCCCAACACATATTTTGTGTGAATATAGAAATTTATTTTATATA 4121
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 450 GGAATAATTA - AGTTAAAGGACAAACATGAAGTTCCTTATGCTATTTTATAGACCTTTG 507
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4120 TGGAAATTAATCTGTTTAAGAAACTGAATAAGGTAATAATTTATTTATTTACTACTATATA 4061
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 508 TAAACAAAGGGGACTTGTGTGAGAAGTCCTGTTTTATACCTTGGAGCAAAACATATCAA 567
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4060 AAATTACAAAGGCTCTTAATGTCGTCGCTTTTACCCTTTAAAAATRAATTTTACAC 4001
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 568 TGTAAAAATAAACAACCTGTTATTTTTTTTTTCTTAAGAGGTA 614
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4000 TAGACGAAATATATTTTAAGTTAAATTTCTAATATTTAGGATTTTA 3954
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 12
US-09-357-206A-20/c
; Sequence 20, Application US/09357206A
; Patent No. 6372962
; GENERAL INFORMATION:
; APPLICANT: Dinesh-Kumar, S.
; APPLICANT: Baker, Barbara
; TITLE OF INVENTION: Pathogen Resistance in Plants using cDNA-N/Intron Constructs
; FILE REFERENCE: 042250/191805 (5830-5)
; CURRENT APPLICATION NUMBER: US/09/357,206A
; CURRENT FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: US 60/093,494
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 6325
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: cDNA-N/Intron construct: E1-I1-E2-I2-E3-I3-E4-E5
US-09-357-206A-20

```

	Query Match	4.7%	Score 37;	DB 4;	Length 6325;
	Best Local Similarity	50.7%;	Prod. No. 0.7;		
	Matches 115;	Conservative 0;	Mismatches 110;	Indels 2;	Gaps 1;
Qy	390	TTGACATTGTTATGCTGCAAGATTAAATGTGGTTTCAGATCTGGGGGTATCTGGTAAACT	449		
Db	4410	TTGAATTACGAGGGCCACACATATTTTGTGCGAATATAGAAATTTATTTTATATATA	4351		
Qy	450	GGATATATTA--AGTTAAAGGACAAACATGAAGTTCCCTTATGTATTTTATAGACCTTTG	507		
Db	4350	TGGAAATATTAGCTTGTTAAGAAACTGAATAAAGGTAAATATTATATTTATATACTATATA	4291		
Qy	508	TAAACAAAGGGGACTCTGTGAGAGTCCCTGTTTTATACCTTGAGCAACATTTACAA	567		
Db	4290	AAATTACAAAGGCTTAAATGTCGTCTTTTACCCCTTAAAAATAAATTTTACAC	4231		
Qy	568	TGTAATAATAACAAACCTGTTATTTTTTTTTTCTTAAAGAGGTAA	614		
Db	4230	TAGACGAAATATATTTTAAAGTCTAAATTTCTAAATATTTAGGATTTTA	4184		

RESULT 13
 US-09-357-206A-22/c
 ; Sequence 22, Application US/09357206A
 ; Patent No. 6372962
 ; GENERAL INFORMATION:
 ; APPLICANT: Dinesh-Kumar, S.
 ; APPLICANT: Baker, Barbara
 ; TITLE OF INVENTION: Pathogen Resistance in Plants using cdNA-N/Intron Constructs
 ; FILE REFERENCE: 042250/191805 (5830-5)
 ; CURRENT APPLICATION NUMBER: US/09/357,206A
 ; CURRENT FILING DATE: 1999-07-20
 ; PRIOR APPLICATION NUMBER: US 60/093,494
 ; PRIOR FILING DATE: 1998-07-20
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 22
 ; LENGTH: 6428
 ; TYPE: DNA
 ; ORGANISM: artificial
 ; FEATURE:
 ; OTHER INFORMATION: cdNA-N/Intron construct: E1-E2-12-E3-I3-E4-I4-E5
 US-09-357-206A-22

 Query Match 4.7%; Score 37; DB 4; Length 6428;
 Best Local Similarity 50.7%; Pred. No. 0.7;
 Matches 115; Conservative 0; Mismatches 110; Indels 2; Gaps 1;

 QY 390 TTGACATTGTATGCTGCAAGATTAAATGCTGTTGCAGATCTGGGGTATCTCGTAAACT 449
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
 Db 4180 TTGATTTACGAGGCCACACATATTTTGTGTGAATATAGAAATTTATTTTTATATA 4121
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
 QY 450 GGAATAATTA--AGTTAAAGGACAAACATGAAGTTCCTTATGTATTTTATAGACCTTG 507
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
 Db 4120 TGGAAAAATCTGTTTAAAGAACTGAATAAAGGCTAAATTTATTTATATATATTA 4061
 |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
 QY 508 TAAACAAAGGGACATGTTGAGAAGTCCTGTTTTTATACCTTGGAGCAAAACATTACAA 567
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 4060 AAATTACAAAGGCTCTTAATGTCTGCTCTTTTACCCTTTTAAANAATAATTTTACAC 4001
 || || || || || || || || || || || || || || || || || || || || || ||
 QY 568 TGTAAAAATAAACAAACCTGTTATTTTCTTTTAAAGAAGTAA 614
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 4000 TAGACGAATATATTAGTTAATTTCTAATATTAGGATTTTA 3954
 || || || || || || || || || || || || || || || || || || || || || ||

 RESULT 14
 US-08-261-663A-1/c
 ; Sequence 1, Application US/08261663A
 ; Patent No. 5571706
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Barbara J
 ; APPLICANT: Whitham, Steven A
 ; TITLE OF INVENTION: Plant Virus Resistance Gene and Methods
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Margaret A. Connor, USDA-ARS
 ; STREET: 800 Buchanan Street
 ; CITY: Albany
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94710
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/261,663A
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Connor, Margaret A
 ; REGISTRATION NUMBER: 30043
 ; REFERENCE/DOCKET NUMBER: 0094.94

```
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/07754A
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Condor, Margaret A
;; REGISTRATION NUMBER: 30043
;; REFERENCE/DOCKET NUMBER: 0094.94
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 559-6067
;; TELEFAX: (510) 559-5777
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 7400 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; ORIGINAL SOURCE:
;; ORGANISM: Nicotiana glauca
;; TISSUE TYPE: leaf
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: join(294..772, 1003..2098, 2941..3213, 5032..6600,
;; LOCATION: 6934..6951)
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 773..1002
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 2099..2940
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 3214..5031
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 6601..6933
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: join(294..772, 1003..2098, 2941..3213, 5032..6600,
;; LOCATION: 6934..6951)
;; PCT-US95-07754A-1
```

```
Query Match 4.7%; Score 37; DB 5; Length 7400;
Best Local Similarity 50.7%; Pred. No. 0.75;
Matches 115; Conservative 0; Mismatches 110; Indels 2; Gaps 1;

Qy 390 TTGACATTGTTATGCTGCAAGATTAAATGTTGTCAGATCTGGGGTATCTGGTAAACT 449
Db 4703 TTGAATTACGAGGGCCACACATTATTTTCTGTAATAGAAATTTATTTTATATA 4644

Qy 450 GGAATAAATA--AGTTAAGGACAAACATGAAGTTCCTATGTATTTTATAGACCTTTG 507
Db 4643 TCGAAATTTACTTGTAAAGAACTGAATAAGGTAATTTATATTTTATATATATA 4584

Qy 508 TAAACAAAAGGGGACTTGTGAGAGTCTCTGTTTTTATACCTTGGAGCAAAACATTACAA 567
Db 4583 AAATTACAAAGGCTCTTAATGTCGTCCTTTTTTACCCCTTTAAATAAATTTTACAC 4524

Qy 568 TGTAATAAATAACAAACCTGTTATTTTCTTAAAGAGGTAA 614
Db 4523 TAGACCAATATATTTTAAGTTAATTTTCTAATATTTAGGATTTTA 4477
```

Search completed: January 17, 2003, 02:02:32
Job time : 100 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 16, 2003, 21:51:23 ; Search time 143 Seconds
(without alignments)
75.478 Million cell updates/sec

Title: US-09-648-310-2
Perfect score: 406
Sequence: 1 MNVEHEVLLVEIHLGSK.....AGELLQGVHDDVDIVLLQD 81

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	406	100.0	81	23	AAU76532
2	390	96.1	81	23	AAU76533
3	390	96.1	142	21	AAB56907
4	375	92.4	83	21	AAB32056
5	302	74.4	87	21	AAG58854
6	286	70.4	83	21	AAB32055
7	286	70.4	95	21	AAG58354
8	286	70.4	161	21	AAG58353
9	232	57.1	92	20	AAV39325
10	158	38.9	77	22	ABBI4740
					Rat Progression Su
					Human Progression
					Human prostate can
					Human secreted pro
					Zea mays protein f
					Human secreted pro
					Arabidopsis thalia
					Arabidopsis thalia
					PSgen13 protein.
					Human nervous syst

11	106	26.1	456	22	ABG27995	Novel human diagno
12	81.5	20.1	183	22	ABB58876	Drosophila melanog
13	73	18.0	399	22	ABB60049	Drosophila melanog
14	66.5	16.4	116	22	AAB46383	H. pylori HPC024 p
15	65	16.0	324	21	AAV91948	Human cytoskeleton
16	65	16.0	407	23	ABB97460	Novel human protei
17	64	15.8	671	23	ABP51430	Human MDP1 SEQ ID
18	63.5	15.6	496	23	ABB92016	Herbicidally activ
19	63	15.5	575	22	AAU50377	Propionibacterium
20	62.5	15.4	491	23	ABB92013	Herbicidally activ
21	62.5	15.4	491	23	ABB92014	Herbicidally activ
22	62.5	15.4	565	19	AAW98798	H. pylori GHPO 128
23	62.5	15.4	565	19	AAW71554	Helicobacter polyp
24	62.5	15.4	565	22	AAB46327	H. pylori HPS024 p
25	62	15.3	137	22	AAO06434	Human polypeptide
26	61.5	15.1	676	23	ABP41814	Human ovarian anti
27	61	15.0	556	20	RAY38731	Neisseria meningit
28	60.5	14.9	101	21	AAV93885	Amino acid sequenc
29	60.5	14.9	387	21	AAG21474	Arabidopsis thalia
30	60.5	14.9	387	21	AAG32491	Arabidopsis thalia
31	60.5	14.9	387	21	AAG49686	Arabidopsis thalia
32	60.5	14.9	398	21	AAG21473	Arabidopsis thalia
33	60.5	14.9	398	21	AAG32490	Arabidopsis thalia
34	60.5	14.9	398	21	AAG49685	Arabidopsis thalia
35	60.5	14.9	573	17	AAW00164	Myosin heavy chain
36	60.5	14.9	573	21	AAG21472	Arabidopsis thalia
37	60.5	14.9	573	21	AAG32489	Arabidopsis thalia
38	60.5	14.9	573	21	AAG49684	Arabidopsis thalia
39	60.5	14.9	798	22	ABB46331	Drosophila melanog
40	60.5	14.9	1113	22	ABB64731	Drosophila melanog
41	60	14.8	1305	21	AAV12875	Murine JNK3 bindin
42	60	14.8	1314	21	AAV12876	Murine JNK3 bindin
43	60	14.8	1336	21	AAV12878	Murine JNK3 bindin
44	60	14.8	1337	21	AAV12877	Murine JNK3 bindin
45	59.5	14.7	157	22	AAO01991	Human polypeptide

ALIGNMENTS

RESULT 1
AAU76532
ID AAU76532 standard; Protein; 81 AA.
XX
AC AAU76532;
XX
DF 05-JUN-2002 (first entry)
XX
DE Rat Progression Suppressed Gene 13 (rPSGen 13).
XX
KW Rat; Progression Suppressed Gene 13; rPSGen 13; cytostatic; cancer;
KW blood vessel; nasopharyngeal tumour; thyroid tumour; leukaemia;
KW lymphoma; breast; lung; prostate; ovary; colon.
XX
OS Rattus sp.
XX
PN WO200216419-A2.
XX
PD 28-FEB-2002.
XX
PF 27-AUG-2001; 2001WO-US26795.
XX
PR 25-AUG-2000; 2000US-0648310.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Fisher PB, Kang D, Su Z;
XX
DR WPI: 2002-280914/32.
DR N-PSDB; ABK11085.
XX
PT New rat and human Progression Suppressed Gene 13 for preventing the growth of cancer cells and/or new blood vessels, and for treating

PT patients suffering from a cancer -
 XX Claim 28; Fig 1; 53pp; English.
 XX
 CC The invention relates to novel isolated nucleic acids which encode a
 CC rat or human Progression Suppressed Gene 13 (PSGen 13) protein. The
 CC nucleic acids are useful for preventing the growth of cancer cells and/or
 CC new blood vessels, and for treating patients suffering from a cancer,
 CC e.g. nasopharyngeal tumour, thyroid tumour, leukaemia, lymphoma, or
 CC cancer of the breast, lung, prostate, ovary or colon. PSGen 13 may also
 CC be used to suppress the transformed phenotype of a malignant cell.
 CC Administration of PSGen 13 gene or protein may result in a decrease in
 CC tumour mass, number of cancer cells, serum tumour marker, tumour
 CC metastasis, vascularisation, perfusion, or rate of tumour growth.
 CC Improved clinical symptoms, and/or increased patient survival. The
 CC present sequence represents the amino acid sequence of rat Progression
 CC Suppressed Gene 13 (rPSGen 13).
 XX
 SQ Sequence 81 AA;
 Query Match 100.0%; Score 406; DB 23; Length 81;
 Best Local Similarity 100.0%; Pred. No. 5.9e-46;
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNVEHEVNLVEEIHRLGSKNADGKLSVKGVLFGVLFQDDRCANLFEALVGTLLKAARRKIYV 60
 DB 1 MNVEHEVNLVEEIHRLGSKNADGKLSVKGVLFGVLFQDDRCANLFEALVGTLLKAARRKIYV 60
 QY 61 YAGELLQGVHDDVDIVLLQD 81
 DB 61 YAGELLQGVHDDVDIVLLQD 81
 RESULT 2
 AAU76533
 ID AAU76533 standard; Protein; 81 AA.
 AC AAU76533;
 DT 05-JUN-2002 (first entry)
 XX Human Progression Suppressed Gene 13 (rPSGen 13).
 DE Human; Progression Suppressed Gene 13; rPSGen 13; cytostatic; cancer;
 KW Blood vessel; nasopharyngeal tumour; thyroid tumour; leukaemia;
 KW lymphoma; breast; lung; prostate; ovary; colon.
 XX Homo sapiens.
 OS
 PN WO200216419-A2.
 XX 28-FEB-2002.
 XX 27-AUG-2001; 2001WO-US26795.
 XX 25-AUG-2000; 2000US-0648310.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX Fisher PB, Kang D, Su Z;
 XX WPI; 2002-280914/32.
 DR N-PSDB; ABK11086.
 XX New rat and human Progression Suppressed Gene 13 for preventing the
 PT growth of cancer cells and/or new blood vessels, and for treating
 PT patients suffering from a cancer -
 XX
 PS Claim 30; Fig 2; 53pp; English.
 XX The invention relates to novel isolated nucleic acids which encode a
 CC rat or human Progression Suppressed Gene 13 (PSGen 13) protein. The
 CC nucleic acids are useful for preventing the growth of cancer cells and/or

CC new blood vessels, and for treating patients suffering from a cancer,
 CC e.g. nasopharyngeal tumour, thyroid tumour, leukaemia, lymphoma, or
 CC cancer of the breast, lung, prostate, ovary or colon. PSGen 13 may also
 CC be used to suppress the transformed phenotype of a malignant cell.
 CC Administration of PSGen 13 gene or protein may result in a decrease in
 CC tumour mass, number of cancer cells, serum tumour marker, tumour
 CC metastasis, vascularisation, perfusion, or rate of tumour growth.
 CC Improved clinical symptoms, and/or increased patient survival. The
 CC present sequence represents the amino acid sequence of human Progression
 CC Suppressed Gene 13 (hPSGen 13).
 XX
 SQ Sequence 81 AA;
 Query Match 96.1%; Score 390; DB 23; Length 81;
 Best Local Similarity 93.8%; Pred. No. 7.7e-44;
 Matches 76; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNVEHEVNLVEEIHRLGSKNADGKLSVKGVLFGVLFQDDRCANLFEALVGTLLKAARRKIYV 60
 DB 1 MNVEHEVNLVEEIHRLGSKNADGKLSVKGVLFGVLFQDDRCANLFEALVGTLLKAARRKIYV 60
 QY 61 YAGELLQGVHDDVDIVLLQD 81
 DB 61 YAGELLQGVHDDVDIVLLQD 81
 RESULT 3
 AAB56907
 ID AAB56907 standard; Protein; 142 AA.
 AC AAB56907;
 XX 13-MAR-2001 (first entry)
 DT Human prostate cancer antigen protein sequence SEQ ID NO:1485.
 DE Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease.
 XX Homo sapiens.
 OS
 PN WO200055174-A1.
 XX 21-SEP-2000.
 XX 08-MAR-2000; 2000WO-US05988.
 XX 12-MAR-1999; 99US-0124270.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX Rosen CA, Ruben SM;
 XX WPI; 2000-587513/55.
 DR N-PSDB; AAF16110.
 XX Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -
 XX
 PS Claim 11; Page 1922; 2338pp; English.
 XX AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
 CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143442.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

CC	allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC	diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC	arthritis and ulcerative colitis; (c) cardiovascular disorders such
CC	as myocardial ischaemias; (d) wound healing; (e) neurological diseases
CC	e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such
CC	as viral, bacterial, fungal and parasitic infections.
XX	
XX	Sequence 83 AA;
XX	Query Match 70.4%; Score 286; DB 21; Length 83;
XX	Best Local Similarity 70.5%; Pred. No. 4.6e-30;
XX	Matches 55; Conservative 8; Mismatches 15; Indels 0; Gaps 0;
QY	1 MNVEHEYNLLVEEIHRLGSKNADGKLSYKFGVLFFQDDRCANLFEALVGTLKAARKRKIVT 60
DB	6 MNVDEEIQKLEEHRLGSRQTDGSKYKTFGVLFNDDRCANIFEALVGTLRAAKRKIVA 65
QY	61 YAGELLQGVHDDVDIVL 78
DB	66 FEGELLQGVHDKVEITL 83
XX	
XX	RESULT 7
XX	AAG58354
XX	ID AAG58354 standard; Protein; 95 AA.
XX	AC AAG58354;
XX	XX
XX	XX
DT	18-OCT-2000 (first entry)
XX	XX
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 75317.
XX	XX
KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	XX
OS	Arabidopsis thaliana.
XX	XX
XX	XX
PN	EP1033405-A2.
XX	XX
PD	06-SEP-2000.
XX	XX
XX	25-FEB-2000; 2000EP-0301439.
XX	XX
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	29-MAR-1999; 99US-0126264.
PR	01-APR-1999; 99US-0126785.
PR	06-APR-1999; 99US-0127462.
PR	08-APR-1999; 99US-0128234.
PR	16-APR-1999; 99US-0128714.
PR	19-APR-1999; 99US-0129845.
PR	21-APR-1999; 99US-0130077.
PR	23-APR-1999; 99US-0130449.
PR	28-APR-1999; 99US-0130510.
PR	30-APR-1999; 99US-0130891.
PR	04-MAY-1999; 99US-0131449.
PR	05-MAY-1999; 99US-0132048.
PR	06-MAY-1999; 99US-0132486.
PR	07-MAY-1999; 99US-0132487.
PR	11-MAY-1999; 99US-0132863.
PR	14-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	18-MAY-1999; 99US-0134370.
PR	19-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.

PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0136029.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137328.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143342.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	19-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.

```
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 70.4%; Score 286; DB 21; Length 95;
Best Local Similarity 70.5%; Pred. No. 5.5e-30;
Matches 55; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

Qy 1 MNVEHEVNLVEEIHRLGSKNADGKLSVKGVLFFODDRCANLFEALVGLKAARKRKIVT 60
   |||: | : | |||||: | | ||||| |||||: |||||: |||||: |||||: |||||
Db 1 MNVDEIQKLEEHRLGSRQDGSYKVTGVLFFNDRCANLFEALVGLRAARKRKIVA 60
   : |||||: | : |
Qy 61 YAGELLQGVHDDIVL 78
   : |||||: | : |
Db 61 FEGELLQGVHDKVEITL 78

RESULT 8
AAG58353
ID AAG58353 standard; Protein; 161 AA.
XX AC AAG58353;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 75316.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
```

PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 Query Match 70.4%; Score 286; DB 21; Length 161;
 Best Local Similarity 70.5%; Pred. NO. 1.le-29;
 Matches 55; Conservative 8; Mismatches 15; Indels 0; Gaps 0;
 QY 1 MNVEHEVNLVEEIHRLGSKNADGKLSVKGVLFODDRCANLFEALVGLTKAARKRIYV 60
 Db 67 MNVDEIQKLEEEIHRLGSRQTDGYSKYTFGLVFNDDRCANIFEALVGLTKAARKRIYA 126
 QY 61 YAGELLQGVHDDVDIVL 78
 Db 127 FEGELLQGVHDKVEITL 144
 RESULT 9
 ID AAY39325 standard; Protein; 92 AA.
 XX AC AAY39325;
 XX DT 01-DEC-1999 (first entry)
 XX DE PSGen13 protein.
 XX KW Progression suppressed gene; PSGen; progression elevated gene; PEGen;
 XX KW tumour; reciprocal subtraction differential RNA display; RSDD;
 XX KW differential expression; gene cloning; cancer.
 XX OS Rattus sp.
 XX PN WO9943844-A1.
 XX PD 02-SEP-1999.
 XX PF 26-FEB-1999; 99WO-US04323.
 XX PR 27-FEB-1998; 98US-0032684.
 XX PR 03-NOV-1998; 98US-0185115.
 XX PR 23-NOV-1998; 98US-0197889.
 XX PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX PI Fisher PB;
 XX DR WPI; 1999-550872/46.
 XX DR N-PSDB; AAZ21517.
 XX PT Identifying nucleic acids differentially expressed between two samples,
 XX PS particularly sequences involved in tumour progression -
 XX PS Claim 40; Fig 35B; 110pp; English.
 This is the amino acid sequence of the PSGen13 protein (progression suppressed gene 13). PSGen13 has suppressed expression in progressed tumour cells. The PSGen13 nucleic acid sequence was identified using new methods for identifying nucleic acids differentially expressed between two samples. The method involves performing reciprocal subtraction differential RNA display (RSDD) between the two samples to generate two subtraction samples. The subtraction samples are amplified and compared to identify those nucleic acids that are differentially expressed. The method is used to identify and clone differentially expressed genes, particularly those with increased or reduced expression during tumour cell progression, e.g. progression suppressed genes (PSGen) and progression elevated genes (PEGen). The method reduces the complexity of the band pattern produced in conventional differential RNA display (where bands may be obscured, resulting in false positive signals) since most bands common to both samples are eliminated, allowing identification and cloning of genes displaying anticipated differential expression. RSDD requires only a single anchored primer for amplification and reamplified cDNA can be analysed by reverse Northern blotting.

```
SQ Sequence 92 AA;
Query Match 57.1%; Score 232; DB 20; Length 92;
Best Local Similarity 72.7%; Pred. No. 7.4e-23;
Matches 48; Conservative 4; Mismatches 8; Indels 6; Gaps 1;

QY 1 MNVEHVNLLVEIHRIGSKNADGKLSVFGVLFQDDRCANLFEALVGLTKAAKRKIYV 60
      |||||
Db 1 MNVEHVNLLVEIHRIGSKNADGKLSVFGVLFQDDRCANLFEALVGLTKAAKRKIYV 60
      |||||
QY 61 YAGELL 66
      |||
Db 57 --GRLL 60

RESULT 10
ABBI14740
ID ABBI14740 standard; Protein; 77 AA.
XX
AC ABBI14740;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polypeptide SEQ ID NO 3397.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0203515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225475.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229387.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
```

17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
(HUMA-) HUMAN GENOME SCI INC.
XX
Rosen CA, Barash SC, Ruben SM;
PI
WPI: 2001-541565/60.
N-PSDB; AB11066.
XX
Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
Claim 11; SEQ ID NO 3397; 1701pp + Sequence Listing; English.
PS
XX The invention relates to novel genes (AB111004-ABA21534) and proteins
XX (AB114678-AB118001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: the sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

QA	SQ	Sequence	77 AA;	Query Match	38.9%;	Score 158;	DB 22;	Length 77;
		Best Local Similarity	96.9%;					
		Matches 31;	Conservative	1;	Mismatches	0;	Indels	0;
QY		1	MNVEHEVNLVVEEIHRLGSKNADGKLSVKFGV	32				
DB		35	MNYDHEVNLVVEEIHRLGSKNADGKLSVKFGV	66				

RESULT 11.

ABG27995	ID	ABG27995 standard; Protein; 456 AA.
XX	AC	ABG27995;
XX	DT	18-FEB-2002 (first entry)
XX	XX	Novel human diagnostic protein #27986.
XX	DE	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	XX	food supplement; medical imaging; diagnostic; genetic disorder.
XX	OS	Homo sapiens.
XX	PN	WO200175067-A2.
XX	PD	11-OCT-2001.
XX	PF	30-MAR-2001; 2001WO-US08631.
XX	PR	31-MAR-2000; 2000US-0540217.
XX	PR	23-AUG-2000; 2000US-0649167.
XX	PA	(HYSE-) HYSEQ INC.
XX	PI	Dmanac RT, Liu C, Tang YT;
XX	DR	WPI; 2001-639362/73.
XX	DR	N-PSDB; AAS92182.
XX	PT	New isolated polynucleotide and encoded polypeptides, useful in
XX	PT	diagnostics, forensics, gene mapping, identification of mutations
XX	PT	responsible for genetic disorders or other traits, and to assess
XX	PT	biodiversity
XX	PS	Claim 20; SEQ ID No 58354; 103pp; English.
XX	CC	The invention relates to isolated polynucleotide (I) and
XX	CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX	CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX	CC	and gene mapping, and in recombinant production of (II). The
XX	CC	polynucleotides are also used in diagnostics as expressed sequence tags
XX	CC	for identifying expressed genes. (I) is useful in gene therapy techniques
XX	CC	to restore normal activity of (II) or to treat disease states involving
XX	CC	(II). (II) is useful for generating antibodies against it, detecting or
XX	CC	quantitating a polypeptide in tissue, as molecular weight markers and as
XX	CC	a food supplement. (II) and its binding partners are useful in medical
XX	CC	imaging of sites expressing (II). (I) and (II) are useful for treating
XX	CC	disorders involving aberrant protein expression or biological activity.
XX	CC	The polypeptide and polynucleotide sequences have applications in
XX	CC	diagnostics, forensics, gene mapping, identification of mutations
XX	CC	responsible for genetic disorders or other traits to assess biodiversity
XX	CC	and to produce other types of data and products dependent on DNA and
XX	CC	amino acid sequences. ABG00010-ABG30377 represent novel human
XX	CC	diagnostic amino acid sequences of the invention.
XX	CC	Note: The sequence data for this patent did not appear in the printed
XX	CC	specification, but was obtained in electronic format directly from WIPO
XX	CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	SQ	Sequence 456 AA:
QY	Query Match	26.1%; Score 106; DB 22; Length 456;
db	Best Local Similarity	43.1%; Pred. No. 2.8e-05;
	Matches 25; Conservative 10; Mismatches 21; Indels 2; Gaps 1;	
	23 DGLSLVKEGVLFODDCRANFEALVGTLKAAKRRIYTYAGELLQGVDHDDIVLIQ 80	
	: : : : : : : :	
	401 DGKIQTTFGLF--DRVYRTSDVKVGTLMARKRGFLVDFEGEMLQGRDHDHVITLK 456	

```

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL04152.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX
XX Disclosure; SEQ ID NO 6939; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 399 AA;
SQ
    Query Match          18.0%; Score 73; DB 22; Length 399;
    Best Local Similarity 40.5%; Pred. No: 0.54;
    Matches 15; Conservative 9; Mismatches 13; Indels 0; Gaps 0;
QY      44 EALYGTLKAARKRKIVTYAGELLGCVHDDVDIVLLQ 80
        :||| |::| |::| |::| |::| |::| |::| |::|
Db      149 DKVVGILLRAKKHLKLVDFEGEMLYQRDRDDVPVFLLK 185
RESULT 14
AAB46383
ID      AAB46383 standard; Protein; 116 AA.
XX
XX AC      AAB46383;
XX
XX DT      05-APR-2001 (first entry)
XX
XX DE      H. pylori HPC024 protein.
XX
XX KW      Microbial infection; antibacterial; Helicobacter pylori infection;
KW vaccine; screening.
XX
XX OS      Helicobacter pylori.
XX
XX WO200073502-A2.
XX
XX PD      07-DEC-2000.
XX
XX PF      31-MAY-2000; 2000WO-EP05024.
XX
XX PR      31-MAY-1999; 99DE-1024965.

```



```

Query Match      16.0%; Score 65; DB 21; Length 324;
Best Local Similarity 29.2%; Pred. No. 4.7;
Matches 21; Conservative 11; Mismatches 30; Indels 10; Gaps 2;

QY       7 VLLVVEIIRLGSKNADGKLSVKFGVLFDQDRCANLF---EALVGTLKAA-----KRR 56
| : :: | ||| :: : | : | |||| : | :
Db       26 VDFIIQGGHLDKAADGNTALHYAALYNQPDCLKLLKGRLVGTVNEAGETALDIARK 85

QY       57 KIVTYAGELLQQ 68
|         ||| |
Db       86 KHHEKEELLEQ 97

```

Search completed: January 17, 2003, 02:06:36
Job time : 144 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 17, 2003, 02:00:53 ; Search time 114 Seconds

(without alignments)
146.402 Million cell updates/sec

Title: US-09-648-310-2

Perfect score: 406

Sequence: 1 MNVEVNLVIEIHLRLSK.....AGELLQGVHDDVDIVLLQD 81

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	403	99.3	81	11 Q9CXU3	Q9CXU3 mus musculus
2	390	96.1	81	4 Q9PIF3	Q9PIF3 homo sapien
3	390	96.1	141	4 Q9P0A1	Q9P0A1 homo sapien
4	286	70.4	161	10 Q81881	Q81881 arabidopsis
5	96.5	23.8	162	5 Q62234	Q62234 caenorhabdi
6	82.5	20.3	371	5 Q46052	Q46052 drosophila
7	81.5	20.1	183	5 Q9V2W6	Q9V2W6 drosophila
8	81.5	20.1	289	5 Q22133	Q22133 caenorhabdi
9	71	17.5	2919	12 Q85431	Q85431 rice stripe
10	69	17.0	318	16 Q98QX3	Q98QX3 mycoplasma
11	68.5	16.9	395	16 Q9PP14	Q9PP14 campylobact
12	67.5	16.6	294	5 Q8STP6	Q8STP6 enccephalito
13	67.5	16.6	447	10 Q9SRM7	Q9SRM7 arabidopsis
14	67.5	16.6	462	10 Q93W32	Q93W32 arabidopsis
15	66	16.3	530	10 Q9SN36	Q9SN36 arabidopsis
16	65	16.0	307	4 Q9NXH7	Q9NXH7 homo sapien

17	65	16.0	407	4 Q9NKK2	Q9NKK2 homo sapien
18	65	16.0	903	4 Q8PDY4	Q8PDY4 homo sapien
19	64.5	15.9	537	2 Q43991	Q43991 acinetobact
20	64	15.8	312	17 Q8ZTY1	Q8ZTY1 pyrobaculum
21	64	15.8	599	16 Q9WZM8	Q9WZM8 thermotoga
22	63.5	15.6	271	5 Q9XZ44	Q9XZ44 lutzomyia l
23	63.5	15.6	496	10 Q9ZQ96	Q9ZQ96 arabidopsis
24	63	15.5	514	5 Q21458	Q21458 caenorhabdi
25	63	15.5	2535	10 Q9S240	Q9S240 arabidopsis
26	62.5	15.4	491	10 Q9ZQ99	Q9ZQ99 arabidopsis
27	62.5	15.4	496	10 Q9ZQ98	Q9ZQ98 arabidopsis
28	62.5	15.4	510	16 Q98BA7	Q98BA7 rhizobium l
29	62.5	15.4	565	16 Q24929	Q24929 helicobacte
30	62.5	15.4	631	2 Q9XB14	Q9XB14 bacillus ce
31	62.5	15.4	986	16 Q92R86	Q92R86 rhizobium m
32	62	15.3	220	5 Q9NG93	Q9NG93 physarum po
33	62	15.3	268	16 Q8YKJ4	Q8YKJ4 anabaena sp
34	62	15.3	608	16 Q8R841	Q8R841 thermoanaer
35	61.5	15.1	491	5 Q9GV11	Q9GV11 ephydatia f
36	61.5	15.1	659	5 Q96069	Q96069 ciona intes
37	61.5	15.1	669	5 Q96068	Q96068 ciona intes
38	61.5	15.1	1010	16 Q8UG04	Q8UG04 agrobacteri
39	61	15.0	217	16 Q8XEW7	Q8XEW7 salmonella
40	61	15.0	556	16 Q9JVG1	Q9JVG1 neisseria m
41	61	15.0	627	16 Q9CMX9	Q9CMX9 pasteurella
42	61	15.0	635	5 Q19056	Q19056 caenorhabdi
43	61	15.0	948	12 Q8QZ10	Q8QZ10 rana tigrin
44	61	15.0	3354	5 Q8T101	Q8T101 bombyx mori
45	60.5	14.9	262	16 Q98LJ1	Q98LJ1 rhizobium l

ALIGNMENTS

RESULT 1

Q9CXU3	Q9CXU3	PRELIMINARY;	PRT;	81 AA.
ID	Q9CXU3			
AC	Q9CXU3			
DT	01-JUN-2001 (TRENBLrel. 17, Created)			
DT	01-JUN-2001 (TRENBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)			
DE	3110003A17Rik protein.			
GN	3110003A17Rik.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=EMRYONIC HEAD;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakawa I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,			
RA	Hayashizaki Y.;			
RA	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK013984; BAB29100.1;			
DR	MCD; MGI:1920362; 3110003A17Rik.			

QY	1	MNVHEVNLLVEETHRLGSKNADGKLSVKFGVLFDQDRCANLFEALVGTLKAAKKRKIVT	60
Db	1	MNVHEVNLLVEETHRLGSKNADGKLSVKFGVLFDQDRCANLFEALVGTLKAAKKRKIVT	60
QY	61	YAGELLQGVDHDDVIILLQD	81
Db	61	YAGELLQGVDHDDVIILLQD	81

RESULT 2
 Q9P1F3 PRELIMINARY; PRT; 81 AA.

ID	Q9P1E3	PRELIMINARY;	PRT;	81 AA.
AC	Q9P1E3;			
DT	01-OCT-2000 (TEMBLrel. 15, Created)			
DT	01-OCT-2000 (TEMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TEMBLrel. 19, Last annotation update)			
DE	PRO2013 (Hypothetical 9.1 kDa protein).			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER;			
RA	Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,			
RA	Liu M., He F.;			
RT	*Functional prediction of the coding sequences of 121 new genes			
RT	dereived by analysis of cDNA clones from human fetal liver.;			
RL	Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BLOOD;			
RA	Strausberg R.;			
RL	Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AF116682; AAF71102.1; -			
DR	EMBL; BC014953; AH14953.1; -			
KW	Hypothetical protein.			
SQ	SEQUENCE 81 AA; 9058 MW; 1FDCE7A1D22B171 CRC64;			

Query Match 96.1%; Score 390; DB 4; Length 81;
 Best Local Similarity 93.8%; Pred.No.3.le-38;
 Matches 76; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY	1	MNVHEVNLLVEETHRLGSKNADGKLSVKFGVLFDQDRCANLFEALVGTLKAAKKRKIVT	60
Db	1	MNVHEVNLLVEETHRLGSKNADGKLSVKFGVLFRDKCANLFEALVGTLKAAKKRKIVT	60
QY	61	YAGELLQGVDHDDVIILLQD	81
Db	61	YPGELLQGVHDHDDVIILLQD	81

RESULT 3
 Q9P0A1 PRELIMINARY; PRT; 141 AA.

ID	Q9P0A1	PRELIMINARY;	PRT;	141 AA.
AC	Q9P0A1;			
DT	01-OCT-2000 (TEMBLrel. 15, Created)			
DT	01-OCT-2000 (TEMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TEMBLrel. 15, Last annotation update)			
DE	HSPC280 (Fragment).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BLOOD;			

[illegible]

OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Negligitoidae; Drosophilidae; Drosophila.
OX	NCHI_TaxID=7227;
RP	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BERKELEY;
RC	MEDLINE=20136006; PubMed=10731132;
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA	Groger R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Stutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA	Barton R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Aril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeson K.Y., Betancur P.V., Berhan B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Donson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA	Fosler C., Gabriellian A.E., Garg N.C., Gelbart W.M., Glasser K.,
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.
RT	"The genome sequence of <i>Drosophila melanogaster</i> ."
RL	Science 287:2185-2195(2000).
DR	ENBL: AE003476; AAF47699.1; --
DR	FLYBase; Fgn0035384; CG2113.
SQ	SEQUENCE 183 AA, 20974 MW; 293A8B0EA9898A61 CRC64;

```
Query Match          20.1%; Score 81.5; DB 5; Length 183;  
Best Local Similarity 31.6%; Pred. No. 0.13;  
Matches 24; Conservative 16; Mismatches 33; Indels 3; Gaps  
2;
```

	QY	6	EVNLLVEEHLRGLSKN-ADGKLSVKFGVLFDODRCANLFEALVGTLLKAARKRKIVTYAGE	64
			: : : : : : : : : : : : : : :	
Db	77	EILQLCDLIQSGRRDPIDGRKVLAFLGOLFETYN--NISDKLATLLGARKEYGVDFSGE	134	
			: : : : : : : : : : : : : : :	
Qy	65	LLQGVDHDDVIDLLQ	80	
		: :		
Db	135	TLFQGRDDETPVRLLR	150	

```
RESULT 8  
Q22133 PRELIMINARY; PRT; 289 AA.  
ID Q22133 AC Q22133;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBRel. 01, Last sequence update)  
DI 01-DEC-2001 (TrEMBRel. 19, Last annotation update)  
DE T04A8.4 protein.  
DN T04A8.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea; Rhabditoidea;
```

```

OC Rhabditiidae; Peloderinae; Caenorhabditis.
RX NCBI_TaxID=6239;
RN [1]
RR SEQUENCE FROM N.A.
RA Palmer S.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RR SEQUENCE FROM N.A.
RX MEDLINE=99059613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998);
DR EMBL: Z35663; CAA84722.1; -.
SQ SEQUENCE 289 AA; 33972 MW; 20DB67F731F03098 CRC64;

Query Match 20.1%; Score 81.5; DB 5; Length 289;
Best Local Similarity 33.7%; Pred. No. 0.23;
Matches 29; Conservative 14; Mismatches 26; Indels 17; Gaps 3;

QY 1 MNVEHEVNLVEEIHRLGSKNADCK---LSVKFGVLFODDRCAANLF-----EALVGTAKAA 53
DB 61 VHVCREILFCETI-----DSNADGEHPKHYVKFKLF-----NIYSPYSOKLVGMLIRA 110
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 54 KRRKIVTYAGELLQGVHDDVIDVL 79
DB 111 RYGLVHFEGEMLYQRQDDEKIITML 136
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 9
Q85431 PRELIMINARY; PRT; 2919 AA.
ID AC Q85431
AC Q85431;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE RNA polymerase.
OS Rice stripe virus (isolate T) (RSV).
OC Viruses; ssRNA negative-strand viruses; Tenuivirus.
OC NCBI_TaxID=36394;
RN [1]
RR SEQUENCE FROM N.A.
RP STRAIN-T.
RC
RX MEDLINE=95088603; PubMed=7996149;
RX Toriyama S., Takahashi M., Sano Y., Shimizu T.;
RT "Nucleotide sequence of RNA 1, the largest genomic segment of rice
RT stripe virus, the prototype of the tenuivirus.";
RL J. Gen. Virol. 75:3569-3579(1994).
DR EMBL: D31879; BAA05677.1; -.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR003323; OTU.
DR Pfam; PF02338; OTU; 1.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
DR PROSITE; PS0802; OTU; 1.
SQ SEQUENCE 2919 AA; 336872 MW; 16FB7BC57FB12FC0 CRC64;

Query Match 17.5%; Score 71; DB 12; Length 2919;
Best Local Similarity 29.2%; Pred. No. 75;
Matches 26; Conservative 15; Mismatches 32; Indels 16; Gaps 4;

```

```

Query Match      17.5%; Score 71; DB 12; Length 2919;
Best Local Similarity 29.2%; Pred. No. 75;
Matches 26; Conservative 15; Mismatches 32; Indels 16; Gaps 4;

QY 1 MNVEHEYNLVVEETHRLGSKNADGKLSV-----KFGVLQDDRCANLFFALVGTAKAA 53
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2465 MNGEIEI---VEELNKLDKGGFESHRLALVERIRVGKLGILGSYTKCOORIELDG----- 2516
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 54 KRRKIVTYAGELLQGVHDDVDI-VLLQD 81
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 2517 EGNKTHRYTGEGIWRGSGFDDSDVCIVQD 2545

RESULT 10
Q98QX3
ID Q98QX3
AC Q98QX3;
PRELIMINARY; PRT; 318 AA.

```

DT 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DE 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
 DE Phosphate acetyltransferase (Phosphotransacetylase) (EC 2.3.1.8).
 GN MYPU_2370.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=21267165; PubMed=11353084;
 RA Chamberlain I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
 RA Moszer I., Dywig K., Wroblewski H., Viari A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis".
 RL Nucleic Acids Res. 29:2145-2153(2001).
 DR EMBL; AL445563; CAC13410.1; -;
 DR MypuList; MYPU_2370; -;
 DR InterPro; IPR002505; PTA.PTB.
 DR Pfam; PF01515; PTA.PTB.1.
 KW Transferase; Acyltransferase; Complete proteome.
 SQ SEQUENCE 318 AA; 35015 MW; ECE864C1AAB7E3EF CRC64;

 Query Match 17.0%; Score 69; DB 16; Length 318;
 Best Local Similarity 36.2%; Pred. No. 7.7;
 Matches 29; Conservative 6; Mismatches 25; Indels 20; Gaps 4;

 QY 6 EVNLLVEIHRGSKNADGKLSVKFGVLFQDDR-----CANLFE-----ALVGTLLKAAK 54
 Db EVNLLVEIHRGSKNADGKLSVKFGVLFQDDR-----CANLFE-----ALVGTLLKAAK 54
 QY 55 RRKIVTYAGELLQGVHDDV 74
 Db TRPF--YAMLLKRGFFDGV 114

 RESULT 11
 Q9PP14
 ID Q9PP14 PRELIMINARY; PRT; 395 AA.
 AC Q9PP14;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Putative integral membrane zinc-metalloprotease.
 GN C30723C.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagels K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrett B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences".
 RL Nature 403:665-668(2000).
 DR EMBL; AL139076; CAB72397.1; -;
 DR MEROPS; M48.008; -;
 DR InterPro; IPR001915; Peptidase_M48.
 DR InterPro; IPR000130; Zn_MTPeptidase.
 DR Pfam; PF01435; Peptidase_M48.1.
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 395 AA; 45319 MW; 47237D2859DEB89E CRC64;

Query Match 16.9%; Score 68.5; DB 16; Length 395;
 Best Local Similarity 26.2%; Pred. No. 12;
 Matches 21; Conservative 19; Mismatches 29; Indels 11; Gaps 3;

 QY 2 NVEHEVNLVEE-----IHRGSKNADGKLSVKFGVLFQDDRCANLFEALVGTLLKAAK 54
 Db 200 NLLKISSLMKQCGFSANGVYVIDASKRDKRLNAYFGGLFKSRVV-LFDTL---LKAALN 255
 QY 55 RRKIVTYAGELLQGVHDDV 74
 Db 256 ERELLAVLGHGELGHFVHKDI 275

 RESULT 12
 Q8STP6
 ID Q8STP6 PRELIMINARY; PRT; 294 AA.
 AC Q8STP6;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical protein ECU09_1300.
 GN ECU09_1300.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI_TaxID=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Genoscope;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RX MEDLINE=21576510; PubMed=11719806;
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
 RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RT Encephalitozoon cuniculi".
 RL Nature 414:450-453(2001).
 DR EMBL; AL590451; CAD27101.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 294 AA; 32949 MW; 01845BED48E5DF2 CRC64;

 Query Match 16.6%; Score 67.5; DB 5; Length 294;
 Best Local Similarity 29.9%; Pred. No. 10;
 Matches 29; Conservative 15; Mismatches 24; Indels 29; Gaps 5;

 QY 3 VEHEVNLVEEIHRL-----GSKNADGKLSVKFGVLFQDDRCANLFEALVGTLLKAAKR-- 55
 Db 123 VFHELRIYSQIHLHLLAVEVHGTRQGD-NINKRCGVLSWDEYIARRAEA----LKAANKIV 177
 QY 56 -----RKIVTYAGELLQGVHDDVDIVLLQD 81
 Db 178 MKSPDPPEGLVRKI-----LIHQELLSNRDIALNKD 208

 RESULT 13
 Q9SRM7
 ID Q9SRM7 PRELIMINARY; PRT; 447 AA.
 AC Q9SRM7;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE T19F11.10 protein.
 DE T19F11.10.
 GN T19F11.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T19F11 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009918; AAF02137.1; -
DR InterPro; IPR000782; BIGH3_fasciclin.
DR Pfam; PF02469; Fasciclin; 1.
SQ SEQUENCE 447 AA; 49236 MW; 66CB1062AF267379 CRC64;

Query Match 16.6%; Score 67.5; DB 10; Length 447;
Best Local Similarity 35.7%; Pred. NO. 18;
Matches 20; Conservative 8; Mismatches 11; Indels 17; Gaps 3;

QY 15 HRLGSKNADGKLSVKGVLFDQDRCANLFEALVGLTKAAKRRKIVTYAGELLQGV 70
Db 372 HKVGAKADG--SVKFG---SGDRSAYLFDPDYI-----DGRISVQGI 410

RESULT 14
Q93W32 PRELIMINARY; PRT; 462 AA.
AC Q93W32;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE AT3G11700/T19F11.10 (Hypothetical 50.8 kDa protein).
GN T19F11.10/AT3G11700.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene T19F11.10/AT3G11700 (GI-6041822).";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]

DR EMBL; AY058879; AAL24265.1; -
DR EMBL; AY056197; AAL07046.1; -
DR InterPro; IPR000782; BIGH3_fasciclin.
DR Pfam; PF02469; Fasciclin; 1.
KW Hypothetical protein.
SQ SEQUENCE 462 AA; 50792 MW; 5BE0BCBF4CDD091 CRC64;

Query Match 16.6%; Score 67.5; DB 10; Length 462;
Best Local Similarity 35.7%; Pred. NO. 19;
Matches 20; Conservative 8; Mismatches 11; Indels 17; Gaps 3;

QY 15 HRLGSKNADGKLSVKGVLFDQDRCANLFEALVGLTKAAKRRKIVTYAGELLQGV 70
Db 372 HKVGAKADG--SVKFG---SGDRSAYLFDPDYI-----DGRISVQGI 410

```

```

RESULT 15
Q9SN36 PRELIMINARY; PRT; 530 AA.
ID Q9SN36;
AC Q9SN36;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ion transporter-like protein (Heavy metal transporter).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RP SEQUENCE FROM N.A.
RA Bevan M., Mueller M.W., Muendlein A., Felber R., Bancroft I.,
RA Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]

RP SEQUENCE FROM N.A.
RA Mueller M.W., Muendlein A., Felber R., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]

RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]

RP SEQUENCE FROM N.A.
RC STRAIN-CV. LANDSBERG;
RA Pittman J.K., Hall J.L., Williams L.E.;
RT "Identification of Nramp5, a heavy metal transporter in Arabidopsis thaliana.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035526; CAB37464.1; -
DR EMBL; AL161549; CAB78881.1; -
DR EMBL; AJ292076; CAC27822.1; -
DR InterPro; IPR001046; Nramp.
DR Pfam; PF01566; Nramp; 1.
DR PRINTS; PR00447; NATRESASSCMP.
DR ProDom; PD001861; Nramp; 1.
DR TIGRFAMS; TIGR01197; nramp; 1.
SQ SEQUENCE 530 AA; 58781 MW; 837E2D22F1BB2E37 CRC64;

Query Match 16.3%; Score 66; DB 10; Length 530;
Best Local Similarity 29.5%; Pred. NO. 33;
Matches 23; Conservative 11; Mismatches 36; Indels 8; Gaps 3;

QY 3 VEHEVNLVEEIIHR---LGSKNADGKLSVKGVLFDQDRCANLFEAL---VGTLLKAAR 55
Db 303 VSEMINLFTVAVPAKFGYGTGKQADSGLVNAVGYLQKYGCGVFPILYINGIGLLAAGQS 362

QY 56 RKIV-TYAGELLQGVHD 72
Db 363 STITGYAGQFINEGFLD 380

Search completed: January 17, 2003, 02:16:33
Job time : 117 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2003, 19:06:08 ; Search time 3174 Seconds
(without alignments)
7151.918 Million cell updates/sec

Title: US-09-648-310-1
Perfect score: 780
Sequence: 1 ggcacgagctctctctgtcc.....gtaataaaaaaaatcatgt 780

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pla.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	780	100.0	780	6	AX456990	Sequence
2	780	100.0	780	6	AX456994	Sequence
3	550.4	70.6	90548	2	AC117858	Rattus no
4	550.4	70.6	185634	2	AC117858	Rattus no
5	482.8	61.9	692	10	AF065991	Mus muscu
6	471.6	60.5	294830	2	AC105605	Rattus no
7	384.2	49.3	176083	2	AC102256	Mus muscu
8	384.2	49.3	189170	2	AC102536	Mus muscu
9	335.2	43.0	835	6	AX456992	Sequence
10	335.2	43.0	835	6	AX456995	Sequence
11	334.2	42.9	65563	2	AC117670	Mus muscu
12	334.2	42.8	170882	2	AC107839	Mus muscu
13	320.4	41.1	876	17	AF116682	Homo sapi
14	283.4	35.3	743	9	BC014953	Homo sapi
15	226.2	28.0	100296	9	AL590308	Human DNA
16	206.6	26.5	552	6	AX396724	Sequence
17	155	19.9	185634	2	AC129055	Rattus no
18	109.6	14.1	516	8	AY087101	Arabidops
19	92.8	11.9	368	6	AX341854	Sequence
20	92.2	11.8	406	6	AX408040	Sequence
21	87.4	11.2	190	6	E25942	Blastocyst
22	87.4	11.2	190	6	E25963	Blastocyst
23	81.4	10.4	464	8	SL293059	S.latifolia
24	78.8	10.1	98124	8	AT11611	Arabidops
25	78.8	10.1	199536	8	ATCHRIV79	Arabidops
26	75	9.6	150446	2	AC097277	Oryza sat
27	53.6	6.9	114041	2	AP001868	Homo sapi
28	53.6	6.9	165913	9	AC079616	Homo sapi
29	53.6	6.9	242184	2	AC015899	Homo sapi
30	48.6	6.2	125544	2	AC095139	Rattus no
31	48.4	6.2	32179	3	AF047663	Caenorhab
32	48.4	6.2	252120	2	AC095803	Rattus no
33	48.4	6.2	299864	2	AC006702	Caenorhab
34	47	6.0	132288	9	AC099848	Homo sapi
35	46.8	6.0	98734	2	PFMAL1P2	Plasmodiu
36	46.4	5.9	15161	6	AX252124	Sequence
37	46.4	5.9	15161	6	AX348889	Sequence
38	45.6	5.8	175416	2	AC120291	Rattus no
39	45	5.8	6956	6	AX348658	Sequence
40	44.6	5.7	12062	1	AE007586	Clostridi
41	44.6	5.7	172390	2	AC096250	Rattus no
42	44.4	5.7	186973	2	AC113755	Rattus no
43	44.4	5.7	271313	2	AC098461	Rattus no
44	44.2	5.7	176671	2	AC115133	Rattus no
45	44.2	5.7	234112	3	PFMALP2	Plasmodiu

ALIGNMENTS

RESULT 1
AX456990

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Rattus sp.
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE

AX456990
Sequence 1 from Patent WO0216419.

780 bp
DNA
linear
PAT 06-JUL-2002

Fisher, P.B., Kang, D.C. and Su, Z.Z.
Progression suppressed gene 13 (psgen 13) and uses thereof

JOURNAL		Patent: WO 0216419-A 1 28-FEB-2002;	
FEATURES		The Trustees of Columbia University in the City of New York (US)	
source	Location/Qualifiers		
	1. .780		
	/organism="Rattus sp."		
	/db_xref="taxon:10118"		
BASE COUNT	223 a 151 c 187 g 219 t		
ORIGIN			
Query Match	100.0%; Score 780; DB 6; Length 780;		
Best Local Similarity	100.0%; Pred. No. 5.5e-196;		
Matches	780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1 GGCACGAGCTCTCTCGTCCCTCCCTCTCCACTGCAGCCTTTCTCTTAGCCCGGAACCA 60		
Db	1 GGCACGAGCTCTCTCGTCCCTCCCTCTCCACTGCAGCCTTTCTCTTAGCCCGGAACCA 60		
Qy	61 CTTCCCTTTCTCTGCTCTCTCTCCCTAGGCGCGGAAGCTGAGTGCAGGGTTCAGACCCA 120		
Db	61 CTTCCCTTTCTCTGCTCTCTCTCCCTAGGCGCGGAAGCTGAGTGCAGGGTTCAGACCCA 120		
Qy	121 CGCGCGCAGCAGCTCTTCAGTGAAGAAAGGAAGCAATCGGAGGGTCAGCAATGAACGTGGA 180		
Db	121 CGCGCGCAGCAGCTCTTCAGTGAAGAAAGGAAGCAATCGGAGGGTCAGCAATGAACGTGGA 180		
Qy	181 GCATGAGGTTAACTCCTCGTGGTGGAGAAATTCATCGTCTGGGTTCCAAAAATCCGATGG 240		
Db	181 GCATGAGGTTAACTCCTCGTGGTGGAGAAATTCATCGTCTGGGTTCCAAAAATCCGATGG 240		
Qy	241 GAAACTGAGTGTGAAGTTTGGGGTCTCTTCCAGAGCAGACAGATGTGCCAATCTCTTTGA 300		
Db	241 GAAACTGAGTGTGAAGTTTGGGGTCTCTTCCAGAGCAGACAGATGTGCCAATCTCTTTGA 300		
Qy	301 AGCGTTGGTGGAACTCTGAAAGCGCGAAAGGAAGATTCGTACGTACGACGAGGAGA 360		
Db	301 AGCGTTGGTGGAACTCTGAAAGCGCGAAAGGAAGATTCGTACGTACGACGAGGAGA 360		
Qy	361 GCTGCTTTTGCAGGCTGTTTCATGATGATGTTGACATTTGTTGCTGCAAGATTAAATGTGG 420		
Db	361 GCTGCTTTTGCAGGCTGTTTCATGATGATGTTGACATTTGTTGCTGCAAGATTAAATGTGG 420		
Qy	421 TTTCACAGATCGGGGGTATCTGGTAACTGGAATAATTAAGTTAAAGACCAACATGAAG 480		
Db	421 TTTCACAGATCGGGGGTATCTGGTAACTGGAATAATTAAGTTAAAGACCAACATGAAG 480		
Qy	481 TTCCCTTATGATATTTTATAGACCTTCTGTAACAAAAGGGGACTGTTGAGAAAGTCCTGTT 540		
Db	481 TTCCCTTATGATATTTTATAGACCTTCTGTAACAAAAGGGGACTGTTGAGAAAGTCCTGTT 540		
Qy	541 TTTATACCTTGGAGCAAAACATTACAATGTAAAAATTAACAAAAACCTGTTATTTTTTTT 600		
Db	541 TTTATACCTTGGAGCAAAACATTACAATGTAAAAATTAACAAAAACCTGTTATTTTTTTT 600		
Qy	601 TCTTAAAGAGTAACTCGGAGACGTAGGCAATAAAATGTTTTTCAGAGGTGCGAAAAAGCT 660		
Db	601 TCTTAAAGAGTAACTCGGAGACGTAGGCAATAAAATGTTTTTCAGAGGTGCGAAAAAGCT 660		
Qy	661 TTTGTTTTTCTTAAACCAATCTTACTGCTCCACACATTGACATCCCTCAAGTGAAGAG 720		
Db	661 TTTGTTTTTCTTAAACCAATCTTACTGCTCCACACATTGACATCCCTCAAGTGAAGAG 720		
Qy	721 CGAACTAAAGACCAACTCGCGTGGAAAAATATATGTTTATGTAATAAAAAAAATCATGT 780		
Db	721 CGAACTAAAGACCAACTCGCGTGGAAAAATATATGTTTATGTAATAAAAAAAATCATGT 780		
RESULT 2			
AX456994			
LOCUS	AX456994	780 bp	DNA linear PAT 06-JUL-2002
DEFINITION	Sequence 5 from Patent WO0216419.		
ACCESSION	AX456994		
VERSION	AX456994.1	GI:21715785	
KEYWORDS			

SOURCE	Rattus sp.
ORGANISM	Rattus sp.
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
JOURNAL	Rattus.
REFERENCE	1
AUTHORS	Fisher, P.B., Kang, D.C. and Su, Z.Z.
TITLE	Progression suppressed gene 13 (psgen 13) and uses thereof
JOURNAL	Patent: WO 0216419-A 5 28-FEB-2002;
FEATURES	The Trustees of Columbia University in the City of New York (US)
SOURCE	Location/Qualifiers
source	1..780
BASE COUNT	/organism="Rattus sp."
ORIGIN	/db_xref="taxon:10118"
	223 a 151 c 187 g 219 t
Query Match	100.0%; Score 780; DB 6; Length 780;
Best Local Similarity	100.0%; Pred. No. 5.5e-196;
Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps	
Qy	1 GGCACGAGCTCTCCTCGTCGCCCTCCCTCTCCACTCGAGCCTTCTCTTAGACCCGAACCA 60
Db	1 GGCACGAGCTCTCCTCGTCGCCCTCCCTCTCCACTCGAGCCTTCTCTTAGACCCGAACCA 60
Qy	61 CTTCCTTTCTTGCTTGTTCCTCCTAGGGCGGGAAGCTGAGTGCAGGGTTTCAGACCCA 120
Db	61 CTTCCTTTCTTGCTTGTTCCTCCTAGGGCGGGAAGCTGAGTGCAGGGTTTCAGACCCA 120
Qy	121 CGCGGCGAGCAGCTCTTCAGTGAAGAAGGAGCAATCGGAGGGTCAGCAATGACGTGGA 180
Db	121 CGCGGCGAGCAGCTCTTCAGTGAAGAAGGAGCAANTCGGAGGGTCAGCAATGACGTGGA 180
Qy	181 GCATGAGGTTAACTCCTGCTGGAGGAAATCATCTCTGGGTTCCAAAAATGCCGATGG 240
Db	181 GCATGAGGTTAACTCCTGCTGGAGGAAATCATCTCTGGGTTCCAAAAATGCCGATGG 240
Qy	241 GAACTGAGTGTGAAGTTTGGGGTCTCTTCCAGAGCAGACAGATGTGCCAATCTCTTTGA 300
Db	241 GAACTGAGTGTGAAGTTTGGGGTCTCTTCCAAAGCAGACAGATGTGCCAATCTCTTTGA 300
Qy	301 AGCGTTGGTGGGAACCTCTGAAAGCCGAAAACGAAAGAAATGTTACGTACGCAGGAGA 360
Db	301 AGCGTTGGTGGGAACCTCTGAAAGCCGAAAACGAAAGAAATGTTACGTACGCAGGAGA 360
Qy	361 GCTGCTTTTGCAGGCTGTCATGATGATGTTGACATTGTAATGCTGCAAGATAATGTGG 420
Db	361 GCTGCTTTTGCAGGCTGTCATGATGATGTTGACATTGTAATGCTGCAAGATAATGTGG 420
Qy	421 TTTCAGATCTGGGGGTATCTGGTAAACTGGAATAAATTAAGTTAAAGGACAACATGAAG 480
Db	421 TTTCAGATCTGGGGGTATCTGGTAAACTGGAATAAATTAAGTTAAAGGACAACATGAAG 480
Qy	481 TTCCTTATGATTTTTATAGACCTTTCTTAACAAAAGGGACCTTGTGTGAGAAGTCTCTGT 540
Db	481 TTCCTTATGATTTTTATAGACCTTTCTTAACAAAAGGGACCTTGTGTGAGAAGTCTCTGT 540
Qy	541 TTTATACCTTGGACCAAAACATTAACAATGTAANAATAAACAAACCCTGTTTTTTTTT 600
Db	541 TTTATACCTTGGACCAAAACATTAACAATGTAANAATAAACAAACCCTGTTTTTTTTT 600
Qy	601 TCCTTAAGAAGGTAATCGGGAGACGTAGGCAATAAATGTTTTTCAGAGGTCGGAAGAAGCT 660
Db	601 TCCTTAAGAAGGTAATCGGGAGACGTAGGCAATAAATGTTTTTCAGAGGTCGGAAGAAGCT 660
Qy	661 TTTGTTTTTCTTAACCAATCTTAGTCTCTGCCACATTTGACACTCCGTCGAAGTGGAAG 720
Db	661 TTTGTTTTTCTTAACCAATCTTAGTCTCTGCCACATTTGACACTCCGTCGAAGTGGAAG 720
Qy	721 CGAACTAAAGCAACCAACGCGGTGGAATAATATATGTTTATATAAAAAAATCATGT 780
Db	721 CGAACTAAAGCAACCAACGCGGTGGAATAATATATGTTTATATAAAAAAATCATGT 780

RESULT 3

AC117858/c

LOCUS

DEFINITION

AC117858 90548 bp DNA linear HTG 18-JUL-2002
Rattus norvegicus clone CH230-344121, *** SEQUENCING IN PROGRESS
***, 44 unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC117858
HTG: HTGS, PHASE1.
Norway rat.

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 90548)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alzbrooks,S.B., Amarantunga,H.C., Are,J.R., Ayalew,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeVila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsone,B., Kelly,S., Khan,O., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwuonu,G.,
Ogarene,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

2 (bases 1 to 90548)
Worley,K.C.
Direct Submission

Submitted (11-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

3 (bases 1 to 90548)
Worley,K.C.
Direct Submission

Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 14, 2002 this sequence version replaced gi:20162682.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVS 344121
Center clone name: CH230-344121
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 51579 bases at least Q40
Consensus quality: 55476 bases at least Q30
Consensus quality: 58229 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1374: contig of 1374 bp in length
1375 1474: gap of unknown length
1475 3073: contig of 1599 bp in length
3074 3173: gap of unknown length
3174 4494: contig of 1321 bp in length
4495 4594: gap of unknown length
4595 6029: contig of 1435 bp in length
6030 6129: gap of unknown length
6130 7189: contig of 1060 bp in length
7190 7289: gap of unknown length
7290 8550: contig of 1261 bp in length
8551 8650: gap of unknown length
8651 10344: contig of 1594 bp in length
10345 10344: gap of unknown length
10345 11679: contig of 1335 bp in length
11680 11779: gap of unknown length
11780 13221: contig of 1442 bp in length
13222 13321: gap of unknown length
13322 14558: contig of 1237 bp in length
14559 14658: gap of unknown length
14659 16268: contig of 1610 bp in length
16269 16368: gap of unknown length
16369 17767: contig of 1399 bp in length
17768 17867: gap of unknown length
17868 18989: contig of 1122 bp in length
18990 19089: gap of unknown length
19090 20904: contig of 1815 bp in length
20905 21004: gap of unknown length
21005 22656: contig of 1652 bp in length
22657 22756: gap of unknown length
22757 24988: contig of 2232 bp in length
24989 25088: gap of unknown length
25089 26704: contig of 1616 bp in length
26705 26804: gap of unknown length
26805 28418: contig of 1614 bp in length
28419 28518: gap of unknown length
28519 30708: contig of 2090 bp in length
30709 32506: contig of 1798 bp in length
32507 32605: gap of unknown length
32607 34405: contig of 1799 bp in length
34406 34505: gap of unknown length
34506 36185: contig of 1679 bp in length
36186 36284: gap of unknown length
36285 37812: contig of 1528 bp in length
37813 37912: gap of unknown length
37913 40252: contig of 2340 bp in length
40253 40352: gap of unknown length
40353 42098: contig of 1746 bp in length
42099 42198: gap of unknown length

```

* 42199 44718: contig of 2520 bp in length
* 44719 44818: gap of unknown length
* 44819 46081: contig of 1263 bp in length
* 46082 46181: gap of unknown length
* 46182 48279: contig of 2098 bp in length
* 48280 48379: gap of unknown length
* 48380 50928: contig of 2449 bp in length
* 50929 52697: contig of 1769 bp in length
* 52698 52797: gap of unknown length
* 52798 55203: contig of 2406 bp in length
* 55204 55303: gap of unknown length
* 55304 57359: contig of 2236 bp in length
* 57360 60302: contig of 2643 bp in length
* 60303 62168: gap of unknown length
* 62169 62268: gap of unknown length
* 62269 65030: contig of 2762 bp in length
* 65031 65130: gap of unknown length
* 65131 68157: contig of 3027 bp in length
* 68158 68257: gap of unknown length
* 68258 71150: contig of 2893 bp in length
* 71151 71250: gap of unknown length
* 71251 73412: contig of 2162 bp in length
* 73413 73512: gap of unknown length
* 73513 75946: contig of 2434 bp in length
* 75947 76046: gap of unknown length
* 76047 78555: contig of 2509 bp in length
* 78556 78655: gap of unknown length
* 78656 80976: contig of 2321 bp in length
* 80977 81076: gap of unknown length
* 81077 84298: contig of 3222 bp in length
* 84299 84398: gap of unknown length
* 84399 87634: contig of 3236 bp in length
* 87635 87734: gap of unknown length
* 87735 90548: contig of 2814 bp in length.
FEATURES             Location/Qualifiers
     source            1..90548
                        /organism="Rattus norvegicus"
                        /db_xref="taxon:10116"
                        /clone="CH230-34421"
BASE COUNT   23767 a 19380 c 17912 g 23833 t 5656 others
ORIGIN
Query Match      70.6%; Score 550.4; DB 2; Length 90548;
Best Local Similarity 99.8%; Pred. No. 6.6e-135;
Matches 551; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 229 AATGCCGATGGAACTGAGTGTGAAGTTGGGGTCCCTCTTCCAAAGACGACAGATGTC 288
DB 76843 AGATGCCGATGGAACTGAGTGTGAAGTTGGGGTCCCTCTTCCAAAGACGACAGATGTC 76784
QY 289 CAATCTTTGAAGCGTTGGTGGAACTCTGAAGCGGAAACGGAAGGAAGATTGTTAC 348
DB 76783 CAATCTTTGAAGCGTTGGTGGAACTCTGAAGCGGAAACGGAAGGAAGATTGTTAC 76724
QY 349 GTACGCGAGGAGCTGCTTTTTCGAAGTGTTCATGATGATGTTGACATGCTATTGCTGCA 408
DB 76723 GTACGCGAGGAGCTGCTTTTTCGAAGTGTTCATGATGATGTTGACATGCTATTGCTGCA 76664
QY 409 AGATTAATCTGGTTTTCAGATCTGGGGGTATCTGTGTAAGTGGAACTGAATAATTAAGTTAAAG 468
DB 76663 AGATTAATCTGGTTTTCAGATCTGGGGGTATCTGTGTAAGTGGAACTGAATAATTAAGTTAAAG 76604
QY 469 ACAACATGAAGTTCCTTATGATTTTATAGACCTTTTGAACAAAGGAGGACTTGTTG 528
DB 76603 ACAACATGAAGTTCCTTATGATTTTATAGACCTTTTGAACAAAGGAGGACTTGTTG 76544
QY 529 AGAAGTCCTTTTATACCTTGGACAAACATTCATATGTAATAATAAACAACACCTG 588
DB 76543 AGAAGTCCTTTTATACCTTGGACAAACATTCATATGTAATAATAAACAACACCTG 76484

```

```

QY 589 TTATTTTTTTTCTTAAGAAGGTAATCGGAGACGTAGGCAATAAAATGTTTTTCAGAGG 648
DB 76483 TTATTTTTTTTCTTAAGAAGGTAATCGGAGACGTAGGCAATAAAATGTTTTTCAGAGG 76424
QY 649 TCGGAAAAGCTTTTGTCTTAAACCATTTCTAGTCTCTGCCACACTTGACACTCCGT 708
DB 76423 TCGGAAAAGCTTTTGTCTTAAACCATTTCTAGTCTCTGCCACACTTGACACTCCGT 76364
QY 709 CAAAGTGAGAGCGCAACTAAAGACCAACTCGGTTGGGAAATATTATGTTTATGTAATAA 768
DB 76363 CAAAGTGAGAGCGCAACTAAAGACCAACTCGGTTGGGAAATATTATGTTTATGTAATAA 76304
QY 769 AAAAAATCATGT 780
DB 76303 AAAAAATCATGT 76292

RESULT 4
AC129055/c
LOCUS              185634 bp      DNA      linear      HTG 25-JUL-2002
DEFINITION        Rattus norvegicus clone CH230-143D2, *** SEQUENCING IN PROGRESS
***, 70 unordered pieces.
AC129055
VERSION           AC129055.1 GI:21954964
KEYWORDS          HTG; HTGS PHASE1.
SOURCE            Rattus norvegicus
ORGANISM          Rattus norvegicus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                  Rattus.
REFERENCE          1 (bases 1 to 185634)
AUTHORS            Muzny, D.M., Adams, C., Adlo-Oduola, B., Ali-osman, F.R., Allen, C.,
                  Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
                  Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonn, D.,
                  Bouck, J., Bowie, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P.,
                  Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
                  Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
                  Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
                  Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
                  Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
                  Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
                  Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
                  Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
                  Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
                  Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
                  Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
                  Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
                  Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
                  Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
                  Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
                  Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
                  Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
                  Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
                  Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
                  Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
                  Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
                  Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, N.,
                  Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, A.,
                  Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G.,
                  Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
                  Peters, M., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
                  Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
                  Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
                  Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Stanley, H., Stone, H.,
                  Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
                  Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,
                  Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
                  Williams, G., Williamson, A., Wleczky, R., Woodson, S., Worley, K.,
                  Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
                  Weinstock, G. and Gibbs, R.
TITLE              Direct Submission
JOURNAL            Unpublished

```

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 185634)

Worley, K.C.

Direct Submission

Submitted (25-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GNPS

Center clone name: CH230-143D2

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 124518 bases at least Q40

Consensus quality: 133171 bases at least Q30

Consensus quality: 139212 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1024: contig of 1024 bp in length
* 1025: contig of unknown length
* 1124: contig of 1050 bp in length
* 2175: gap of unknown length
* 2275: contig of 1022 bp in length
* 3297: gap of unknown length
* 3397: contig of 1014 bp in length
* 4411: gap of unknown length
* 4511: contig of 1022 bp in length
* 5532: gap of unknown length
* 5632: contig of 1000 bp in length
* 6732: gap of unknown length
* 7799: contig of 1067 bp in length
* 7899: gap of unknown length
* 8966: contig of 1067 bp in length
* 9067: gap of unknown length
* 9067: contig of 1034 bp in length
* 10101: gap of unknown length
* 10201: contig of 1697 bp in length
* 11898: gap of unknown length
* 11998: contig of 1233 bp in length
* 13231: gap of unknown length
* 13331: contig of 1453 bp in length
* 14783: gap of unknown length
* 14883: contig of 1320 bp in length
* 16204: gap of unknown length
* 16303: contig of 1050 bp in length
* 17354: gap of unknown length
* 17354: contig of 1331 bp in length
* 17454: gap of unknown length
* 18785: contig of 1125 bp in length
* 18885: gap of unknown length
* 20009: contig of 1125 bp in length
* 20109: gap of unknown length
* 20110: contig of 1256 bp in length
* 21366: gap of unknown length
* 21466: contig of 1961 bp in length
* 23427: gap of unknown length
* 23527: contig of 1262 bp in length
* 24789: gap of unknown length
* 24889: contig of 2253 bp in length
* 27142: gap of unknown length
* 27242: contig of 2102 bp in length

29344: gap of unknown length
29444: contig of 1268 bp in length
30811: gap of unknown length
30812: contig of 1214 bp in length
32026: gap of unknown length
32126: contig of 1633 bp in length
33759: gap of unknown length
33859: contig of 1827 bp in length
35686: gap of unknown length
35786: contig of 1511 bp in length
37297: gap of unknown length
37397: contig of 1216 bp in length
38613: gap of unknown length
38713: contig of 1470 bp in length
40183: gap of unknown length
40283: contig of 1416 bp in length
41699: gap of unknown length
41799: contig of 2228 bp in length
44027: gap of unknown length
44127: contig of 1840 bp in length
45967: gap of unknown length
46067: contig of 1057 bp in length
47124: gap of unknown length
47224: contig of 1778 bp in length
49002: gap of unknown length
49102: contig of 1855 bp in length
50957: gap of unknown length
51057: contig of 2173 bp in length
53230: gap of unknown length
53330: contig of 2296 bp in length
55626: gap of unknown length
55726: contig of 2583 bp in length
58309: gap of unknown length
58409: contig of 1481 bp in length
59890: gap of unknown length
59990: contig of 1984 bp in length
61974: gap of unknown length
62074: contig of 2131 bp in length
64205: gap of unknown length
64305: contig of 2521 bp in length
66826: gap of unknown length
66926: contig of 1791 bp in length
68717: gap of unknown length
68817: contig of 2102 bp in length
70919: gap of unknown length
71019: contig of 1920 bp in length
72939: gap of unknown length
73039: contig of 2005 bp in length
75044: gap of unknown length
75144: contig of 2620 bp in length
77764: gap of unknown length
77864: contig of 2488 bp in length
80352: gap of unknown length
80452: contig of 3776 bp in length
84228: gap of unknown length
84328: contig of 2205 bp in length
86533: gap of unknown length
86633: contig of 3360 bp in length
89993: gap of unknown length
90093: contig of 3153 bp in length
93246: gap of unknown length
93346: contig of 3552 bp in length
96898: gap of unknown length
96998: contig of 2363 bp in length
99361: gap of unknown length
99461: contig of 2588 bp in length
102049: gap of unknown length
102149: contig of 3541 bp in length
105690: gap of unknown length
105790: contig of 3109 bp in length
108899: gap of unknown length
112575: contig of 3577 bp in length
112576: gap of unknown length

Query Match 70.6%; Score 550.4; DB 2; Length 185634;
Best Local Similarity 99.8%; Pred. No. 7e-135; Indels 0; Gaps 0;
Matches 551; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 229 AATGCGGATGGGAACTGAGTGTGAAGTTTGGGGTCTCTTCCACAGACACAGATGTC 288
Db 134350 AGATGCGGATGGGAACTGAGTGTGAAGTTTGGGGTCTCTTCCACAGACACAGATGTC 134291
QY 289 CAATCTCTTTGAAGCCTTGGTGGAACTCTGAAGCCGCAAAACGGAAGGAGATGTTAC 348
Db 134290 CAATCTCTTTGAAGCCTTGGTGGAACTCTGAAGCCGCAAAACGGAAGGAGATGTTAC 134231
QY 349 GTACGAGGAGAGCTCTTTTCAAGGTGTTTCATGATGTTGACATGTTGACATGTTGCTGCA 408
Db 134230 GTACGAGGAGAGCTCTTTTCAAGGTGTTTCATGATGTTGACATGTTGACATGTTGCTGCA 134171
QY 409 AGATTAAATGTTTTCAGATCTGGGGGTATCTGGAAGTAACTGGAATAAATTAAGTTAAAG 468
Db 134170 AGATTAAATGTTTTCAGATCTGGGGGTATCTGGAAGTAACTGGAATAAATTAAGTTAAAG 134111
QY 469 ACAACATGAAGTCTCTTATGATTTTATAGACCTTGTAAACAAAGGGACTTGTG 528
Db 134110 ACAACATGAAGTCTCTTATGATTTTATAGACCTTGTAAACAAAGGGACTTGTG 134051
QY 529 AGAAGTCCCTGTTTTTATACCTTGGAGCAAAACATTAACAATGTAAAAATAAACAACCTG 588
Db 134050 AGAAGTCCCTGTTTTTATACCTTGGAGCAAAACATTAACAATGTAAAAATAAACAACCTG 133991
QY 589 TTATTTTTTTTTTCTTAAGAAGTAATCGGGAGACGTAGGCAATAAATGTTTTCAGAGG 648
Db 133990 TTATTTTTTTTTTCTTAAGAAGTAATCGGGAGACGTAGGCAATAAATGTTTTCAGAGG 133931
QY 649 TCGGAAAGCTTTCTTCTTAACACCTCTAGTCTCTGCCACACTTGACACTCCGT 708
Db 133930 TCGGAAAGCTTTCTTCTTAACACCTCTAGTCTCTGCCACACTTGACACTCCGT 133871
QY 709 CAAAGTCAGAGCGAACTAAAGACCACTCGCGTGGGAAATATATGTTTATGTAATAA 768
Db 133870 CAAAGTCAGAGCGAACTAAAGACCACTCGCGTGGGAAATATATGTTTATGTAATAA 133811
QY 769 AAAAAATCATGT 780
Db 133810 AAAAAATCATGT 133799

RESULT 5
AF065991/c
LOCUS AF065991 692 bp mRNA linear ROD 20-OCT-2001
DEFINITION Mus musculus unknown mRNA.
ACCESSION AF065991
VERSION AF065991.1 GI:16303306
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 692)
Ganguly,K., Yang,L.F. and Reddy,P.K.
TITLE A novel cDNA clone from mouse thymus cDNA library
JOURNAL Unpublished
REFERENCE
AUTHORS Ganguly,K., Yang,L.F. and Reddy,P.K.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-1998) Anatomy & Cell Biology, SUNY HSC at
Brooklyn, 450 Clarkson Avenue, Brooklyn, NY 11203, USA
FEATURES
Source
1. .692
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="thymus"
<1. .240
/codon_start=1
CDS

/product="unknown"
/protein_id="AAL16773.1"
/db_xref="GI:16303307"
/translation="EFHFFYSINNNIPMAAGLOFASCFDRMSSGCRNCLNGLREQKL
FHSKNVSLPTFFNYLLKNNRFLCLFIVMFCFKV"
BASE COUNT 199 a 159 c 108 g 226 t
ORIGIN
Query Match 61.9%; Score 482.8; DB 10; Length 692;
Best Local Similarity 85.8%; Pred. No. 3.8e-117; Indels 15; Gaps 4;
Matches 587; Conservative 0; Mismatches 82; Indels 15; Gaps 4;
QY 98 GCTCAGTCCAGGGTTCAGACCCAGCGGAGCAGCTCTTCAGTGAAGAAGCAACAATC 157
Db 685 GCCAGTACAGGGTTCAGACCCAGCGGAGCAGCTCTTCAGTGAAGAAGCAACAATC 627
QY 158 GGAGGGTCAGCAATGACGTGGAGCATGAGGTTAACTCCCTGGTGGAGGAATTCATCGT 217
Db 626 GGAGGGTCAGCAATGACGTGGAGCATGAGGTTAACTCCCTGGTGGAGGAATTCATCGC 567
QY 218 CTGGGTTCCAAAAATCGCGATGGGAACTGAGTGTGAAGTTTGGGGTCTCTTCCAAGAC 277
Db 566 CTGGGTTCCAAAAATCGCGATGGGAACTGAGTGTGAAGTTTGGGGTCTCTTCCAAGAT 507
QY 278 GACAGATGTCGAATCTCTTTGAAGCGTTGGTGGGAACCTCTGAAAGCCGCAAAACGAAAG 337
Db 506 GACAGATGTCGAATCTCTTTGAAGCGTTGGTGGGAACCTCTGAAAGCTGCAAAACGAAAG 447
QY 338 AAGATTGTTACGTACGAGGAGAGCTGCTTTTGAAGGTGTTCAATGATGATGTTGACATT 397
Db 446 AAGATTGTTACGTACGAGGAGAGCTGCTTTTGAAGGTGTTCAATGATGATGTTGACATT 387
QY 398 GTATTGTCGAAGATTAATGTTGTTGCAGATCTGGGGTATCTGTTAACTGGAATAT 457
Db 386 GTATTGTCGAAGATTAATGTTGTTGCAGATCTGGGGTATCTGTTAACTGGAATAT 327
QY 458 TAAGTTAAAGCAAAACATGAAGTTCCTTATG -TATTTTATAGACCTTTGTAACAAA 515
Db 326 TAAGTTAAAGCAAAACATGAAGTTCCTTATGATATATTTTATAGACCTTTGTAACAAA 267
QY 516 AGGGACTGTTGAGAGTCTCTGTTTTATACCTGGAGCAAAACATTAATGTAATAA 575
Db 266 AGGGACTGTTGAGAGTCTCTGTTTTATACCTGGAGCAAAACATTAATGTAATAA 207
QY 576 TAAACAAACCTGTTATTTTTTTTCTTAAGAGGTAATCGGGAGAGCTAGGCAATAA 635
Db 206 TAAACAAACCTGTTATTTTTTTTCTTAAGAGGTAATCGGGAGAGCTAGGCAATAA 153
QY 636 ATGTTTTCAGAGGTCGCAAAAGCTTTTGTCTTCTTAAACCAATTCCTT-----AGTCTCT 689
Db 152 ACATTTTGGAGGTGTGAAAAAGCTTTTGTCTTCTTAAACCAATTCCTTAAAGACAATTTCTA 93
QY 690 GCCACATTCACATCTCCGTCAAGTGAGAGGAACTAAAGCAACCTGCGGTGGAAAT 749
Db 92 CAGGCACTTCACATCTCTGCAAGCAAGCAAACTGCAGACCAGCTGCCATGAAATG 33
QY 750 ATTATGTTTATGTAATAAAAAA 773
Db 32 TTGTTGTTTATGTAATAAAAAA 9
RESULT 6
AC105605
LOCUS AC105605 294830 bp DNA linear HTG 17-JUL-2002
DEFINITION Rattus norvegicus clone CH230-107J10, *** SEQUENCING IN PROGRESS
***, 108 unordered pieces.
ACCESSION AC105605
VERSION AC105605.2 GI:21743919
KEYWORDS HTG; HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE AUTHORS

Rattus
1 (bases 1 to 294830)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barabara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozador,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maneshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G., Orgunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,E., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 294830)
Worley,K.C.
Direct Submission
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 294830)
Worley,K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:18092828.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNHM
Center clone name: CH230-107J10
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 199474 bases at least Q40
Consensus quality: 210759 bases at least Q30
Consensus quality: 218971 bases at least Q20

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

1158: contig of 1158 bp in length
1159: gap of unknown length
1259: contig of 1029 bp in length
2288: 2387: gap of unknown length
2388: 3421: contig of 1034 bp in length
3422: 3521: gap of unknown length
3522: 4584: contig of 1043 bp in length
4565: 4664: gap of unknown length
4665: 5718: contig of 1054 bp in length
5719: 5818: gap of unknown length
5819: 6870: contig of 1052 bp in length
6871: 8115: contig of 1145 bp in length
8116: 8215: gap of unknown length
8216: 9465: contig of 1250 bp in length
9466: 10642: contig of 1077 bp in length
10643: 10742: gap of unknown length
10743: 12141: contig of 1399 bp in length
12142: 12241: gap of unknown length
12242: 13296: contig of 1055 bp in length
13297: 13396: gap of unknown length
13397: 14464: contig of 1068 bp in length
14465: 14584: gap of unknown length
14585: 15885: contig of 1321 bp in length
15886: 15985: gap of unknown length
15986: 17010: contig of 1025 bp in length
17011: 17110: gap of unknown length
17112: 18113: contig of 1003 bp in length
18114: 18213: gap of unknown length
18214: 19538: contig of 1325 bp in length
19539: 19638: gap of unknown length
19639: 21190: contig of 1552 bp in length
21191: 21290: gap of unknown length
21291: 22651: contig of 1361 bp in length
22652: 22751: gap of unknown length
22752: 24131: contig of 1380 bp in length
24132: 24231: gap of unknown length
24232: 25582: contig of 1351 bp in length
25583: 25682: gap of unknown length
25683: 26874: contig of 1192 bp in length
26875: 26974: gap of unknown length
26975: 28108: contig of 1134 bp in length
28109: 28208: gap of unknown length
28209: 29747: contig of 1539 bp in length
29748: 29847: gap of unknown length
29848: 30907: contig of 1060 bp in length
30908: 31007: gap of unknown length
31008: 32166: contig of 1159 bp in length
32167: 32266: gap of unknown length
32267: 33695: contig of 1429 bp in length
33696: 33795: gap of unknown length
33796: 35086: contig of 1291 bp in length
35087: 35186: gap of unknown length
35187: 36654: contig of 1467 bp in length
36654: 36753: gap of unknown length
36754: 37758: contig of 1005 bp in length
37759: 39133: contig of 1275 bp in length
39134: 39233: gap of unknown length
39234: 40301: contig of 1068 bp in length
40302: 40401: gap of unknown length
40402: 41649: contig of 1248 bp in length

REFERENCE AUTHORS TITLE JOURNAL

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 108 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1158: contig of 1158 bp in length
1159: gap of unknown length
1259: contig of 1029 bp in length
2288: 2387: gap of unknown length
2388: 3421: contig of 1034 bp in length
3422: 3521: gap of unknown length
3522: 4584: contig of 1043 bp in length
4565: 4664: gap of unknown length
4665: 5718: contig of 1054 bp in length
5719: 5818: gap of unknown length
5819: 6870: contig of 1052 bp in length
6871: 8115: contig of 1145 bp in length
8116: 8215: gap of unknown length
8216: 9465: contig of 1250 bp in length
9466: 10642: contig of 1077 bp in length
10643: 10742: gap of unknown length
10743: 12141: contig of 1399 bp in length
12142: 12241: gap of unknown length
12242: 13296: contig of 1055 bp in length
13297: 13396: gap of unknown length
13397: 14464: contig of 1068 bp in length
14465: 14584: gap of unknown length
14585: 15885: contig of 1321 bp in length
15886: 15985: gap of unknown length
15986: 17010: contig of 1025 bp in length
17011: 17110: gap of unknown length
17112: 18113: contig of 1003 bp in length
18114: 18213: gap of unknown length
18214: 19538: contig of 1325 bp in length
19539: 19638: gap of unknown length
19639: 21190: contig of 1552 bp in length
21191: 21290: gap of unknown length
21291: 22651: contig of 1361 bp in length
22652: 22751: gap of unknown length
22752: 24131: contig of 1380 bp in length
24132: 24231: gap of unknown length
24232: 25582: contig of 1351 bp in length
25583: 25682: gap of unknown length
25683: 26874: contig of 1192 bp in length
26875: 26974: gap of unknown length
26975: 28108: contig of 1134 bp in length
28109: 28208: gap of unknown length
28209: 29747: contig of 1539 bp in length
29748: 29847: gap of unknown length
29848: 30907: contig of 1060 bp in length
30908: 31007: gap of unknown length
31008: 32166: contig of 1159 bp in length
32167: 32266: gap of unknown length
32267: 33695: contig of 1429 bp in length
33696: 33795: gap of unknown length
33796: 35086: contig of 1291 bp in length
35087: 35186: gap of unknown length
35187: 36654: contig of 1467 bp in length
36654: 36753: gap of unknown length
36754: 37758: contig of 1005 bp in length
37759: 39133: contig of 1275 bp in length
39134: 39233: gap of unknown length
39234: 40301: contig of 1068 bp in length
40302: 40401: gap of unknown length
40402: 41649: contig of 1248 bp in length

COMMENT

Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

COMMENT

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17061342.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L18269

Center clone name: 191_F_12

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 171878 bases at least 840

Consensus quality: 173759 bases at least Q30

Consensus quality: 174226 bases at least Q20

Insert size: 167000; agarose-fp

Insert size: 174483; sum-of-contigs

Quality coverage: 7.3 in Q20 bases; agarose-fp

Quality coverage: 7.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      119: contig of 119 bp in length
120    219: gap of 100 bp
220    1584: contig of 1365 bp in length
1585  1684: gap of 100 bp
1685    44019: contig of 42335 bp in length
44020  44119: gap of 100 bp
44120    47141: contig of 3022 bp in length
47142  47241: gap of 100 bp
47242    51149: contig of 3908 bp in length
51150  51249: gap of 100 bp
51250    55244: contig of 3995 bp in length
55245  55344: gap of 100 bp
55345    57799: contig of 2455 bp in length
57800  57899: gap of 100 bp
57900    61584: contig of 3685 bp in length
61585  61684: gap of 100 bp
61685    66044: contig of 4360 bp in length
66045  66144: gap of 100 bp
66145    73488: contig of 7344 bp in length
73489  73588: gap of 100 bp
73589    83700: contig of 10112 bp in length
83701  83800: gap of 100 bp
83801    95087: contig of 11287 bp in length
95088  95187: gap of 100 bp
95188    109668: contig of 14481 bp in length
109669 109768: gap of 100 bp
109769 128857: contig of 19089 bp in length
128858 128957: gap of 100 bp
128958    151848: contig of 22891 bp in length
151849 151948: gap of 100 bp
151949    175262: contig of 23314 bp in length
175263 175362: gap of 100 bp
175363    176083: contig of 721 bp in length.
```

FEATURES source

```

Location/Qualifiers
1. .176083
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-191F12"
/clone_lib="RPCI-24 Male Mouse BAC"
1. .119
/note="assembly_fragment"
clone_end:SP6
vector_side:left
220. .1584
/note="assembly_fragment"
1685. .44019
/note="assembly_fragment"
44120. .47141
/note="assembly_fragment"
47242. .51149
/note="assembly_fragment"
51250. .55244
/note="assembly_fragment"
55345. .57799
/note="assembly_fragment"
57900. .61584
/note="assembly_fragment"
61685. .66044
/note="assembly_fragment"
66145. .73488
/note="assembly_fragment"
73589. .83700
/note="assembly_fragment"
83801. .95087
/note="assembly_fragment"
95188. .109668
/note="assembly_fragment"
109769. .128857
/note="assembly_fragment"
128958. .151848
/note="assembly_fragment"
151949. .175262
/note="assembly_fragment"
175363. .176083
clone_end:T7
vector_side:right"

BASE COUNT  49418 a 35961 c 36427 g 52675 t 1602 others
ORIGIN
```

```

Query Match      49.3%; Score 384.2; DB 2; Length 176083;
Best Local Similarity 85.1%; Pred. No. 8.7e-91;
Matches 469; Conservative 0; Mismatches 68; Indels 14; Gaps 3;
```

```

QY  229 AAATGCCGATGGAACTGAGTGTGAAGTTGGGGTCCCTTCCAGACGACAGATGTC 288
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164344 AGATGCTGATGGAATAAAGTGAAGTTGGGGTCCCTTCCAGATGACAGATGTC 164403

QY  289 CAATCTCTTTGAAGCGTTGGTGGGAACCTCTGAAGCGGCAAAAGGAGGATTGTTAC 348
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164404 CAATCTCTTTGAAGCGTTGGTAGGAACCTCTGAAGCGTGAAGGAGGATTGTTAC 164463

QY  349 GTACCGAGGAGAGCTGCTTTTTCGAAGTGTTTCATGATGTTGACATTGATTGCTGCA 408
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164464 ATACGCGAGGGAACTACTTTTTCGAAGTGTTTCATGATGATTGACATTGATTGCTGCA 164523

QY  409 AGATTAACTGTTTTCAGATCTGGGGGTATCTGGTAAACTTGAATAAATAAGTTAAAGG 468
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164524 AGATTAACTGTTTTCATGCTGGCTTGGTGTATCTGATAAACTGCAATAAATAAGTTAAAG 164583

QY  469 ACAAAACATCAAGTCTCTTATGATTTTATAGACCTTTGTAAACAAAGGGGACTTTGTC 528
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164584 ACTAAGCGTAATTTCTCTTAATTTTATAGAACTTTGTAAACAAAGGGGCTTTGTTG 164643

QY  529 AGAAGTCCTGTTTATATACCTTGGAGCAAAACATTAACAATGATAAAATAAACAACCTG 588
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

Db 164544 AGAAGTCTGTTTATACCTTGAAGCAAAACATTACAAATGTAATAATAACAAACAC--- 164700
Qy 589 TTATTTTCTTTTCTTAAAGAGTAATCGGAGACGTAGGCAATAAATAATGTTTCAGAGG 648
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164701 ---TATTATTTTCTTAAAGAGTAATGCGAATGAGTAAAGAACATTTTGGAGG 164757
Qy 649 TCGGAAAGCTTTGTTTCTTAAACATTCCTT-----AGTCTCTGCCACACTTGACA 702
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164758 TGTGAAAAGCTTTTGTCTTCTTAAACCATCTTAAAGACAATTTCTCAGGCATTTGACA 164817
Qy 703 CTCGCTCAAGTGAAGCGACGTAAGACCACTCGGTCGGAATAATATTTATGTTATGCT 762
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164818 TTTCTGTAAGACGAAGCAAGCAACTGCAGACCACTGCGCATGAAAATGTT--TGTATG 164875
Qy 763 AATAAAAAAA 773
  ||||| |||||
Db 164876 AATAAAAAAA 164886

RESULT 8
AC102536
LOCUS
DEFINITION
  AC102536 189170 bp DNA linear HTG 21-AUG-2002
MUS musculus clone RP23-69K22, WORKING DRAFT SEQUENCE, 31 unordered
  pieces.
ACCESSION
  AC102536
VERSION
  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
  house mouse.
SOURCE
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 189170)
  Birren,B., Nusbaum,C. and Lander,E.
  Mus musculus, clone RP23-69K22
  Unpublished
  2 (bases 1 to 189170)
  Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
  Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
  Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
  Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
  Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
  Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
  Glnde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
  Hagos,C., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
  Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K.,
  Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
  MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
  McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
  Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
  Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
  Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
  Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
  Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
  Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
  Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
  Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
  Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
  Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  3 (bases 1 to 189170)
  Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
  Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
  Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
  Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
  Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
  Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
  Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
  Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
  Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
  McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
  Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,

```

TITLE JOURNAL COMMENT

O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17061622.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18974
Center clone name: 69_K22
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181583 bases at least Q40
Consensus quality: 184245 bases at least Q30
Consensus quality: 185286 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 186170; sum-of-contigs
Quality coverage: 6.8 in Q20 bases; agarose-fp
Quality coverage: 7.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 965: contig of 965 bp in length
* 1066 1065: gap of 100 bp
* 1066 1065: gap of 100 bp
* 1723 1822: contig of 657 bp in length
* 1823 2591: contig of 769 bp in length
* 2592 2691: gap of 100 bp
* 2692 2691: gap of 100 bp
* 3853 3952: contig of 1161 bp in length
* 3953 3952: gap of 100 bp
* 3953 3952: contig of 1571 bp in length
* 5524 5623: gap of 100 bp
* 5624 6928: contig of 1305 bp in length
* 6929 7028: gap of 100 bp
* 7029 8370: contig of 1342 bp in length
* 8371 8470: gap of 100 bp
* 8471 9644: contig of 1174 bp in length
* 9645 9744: gap of 100 bp
* 9745 11173: contig of 1429 bp in length
* 11174 11273: gap of 100 bp
* 11274 12891: contig of 1618 bp in length
* 12892 12991: gap of 100 bp
* 12992 14773: contig of 1782 bp in length
* 14774 14873: gap of 100 bp
* 14874 16304: contig of 1431 bp in length
* 16305 16404: gap of 100 bp
* 16405 18574: contig of 2170 bp in length
* 18575 19674: gap of 100 bp
* 18675 20391: contig of 1717 bp in length
* 20392 20491: gap of 100 bp
* 20492 23104: contig of 2613 bp in length
* 23105 23204: gap of 100 bp
* 23205 24923: contig of 1719 bp in length

Matches 539; Conservative 0; Mismatches 218; Indels 23; Gaps 5;

QY 16 GGTCCCTCCCTCTCCACTGACGCTTCTCTTAGCCGACCACTTCTCTTCTTGCT 75
 DB 43 CTTGCTCTCTCTCCACAGCGCTTCTCTTCTTCCGACGCACTTCTCTTCTGCGC 102

QY 76 TGTTCCTCCTTAGGGCGGGAAGCTGAGTGCAGGGTTCAGACCCGCGGAGCAGCTC 135
 DB 103 TCACCTCCCTCCAGTGCAGTGAAGAGTAAACCGGTCAGACCCGCGGAGCTTCT 162

QY 136 TTCAGTGAAGAGGAAGCAATCGGAGGCTGAGCAATGAACCTGAGGATGAGTTAACT 195
 DB 163 CCGCGGGAAGGAACCGCCGACAGAGGAGCAATGATGATGATGATGATGATGAT 222

QY 196 CTTGAGGCGGGAAGGAAGTGTGATGATGATGATGATGATGATGATGATGATGAT 255
 DB 343 TCTTAAAGCTGCAAAACGAAGAGATGTAACATATCCAGGAGAGCTGCTTCTGCAAGG 402

QY 376 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 435
 DB 403 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 462

QY 436 GTA-----TCTGTAACTGGAATAATGATGATGATGATGATGATGATGATGAT 477
 DB 463 GTACTGCCATTTTCTGTAAGAGAGTAACTGGAAT-ATAAGTGAAGAGCAAACTTGA 521

QY 478 AGTTTCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 535
 DB 522 ACATACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 581

QY 536 CTGTTTTTATACCTTGGAGCAAAACATTAATGATGATGATGATGATGATGATGATGAT 595
 DB 582 CTGCTCTTTTATATCTTGAAGAGAAATCTATGATGATGATGATGATGATGATGATGAT 641

QY 596 TTTTCTTGAAGAGTAACTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 655
 DB 642 ATTTTCTCAGGAATCTGGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 701

QY 656 AAGCTTTTCTTAAACCACTTCTAGTCT--CTGCCACACTTGACACTCCGTCGAAAG 713
 DB 702 GGAATGTTGTTTATATCTTGAAGAGAAATCTATGATGATGATGATGATGATGATGATGAT 761

QY 714 TGAGAAGGAAGTAAAGCAAACTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 773
 DB 762 CAACAAGCAAACTGAAGCAAACTCCTATGAGAAATATATGATGATGATGATGATGATGATGAT 821

RESULT 10
 LOCUS AX456995 835 bp DNA linear PAT 06-JUL-2002
 DEFINITION Sequence 6 from Patent WO216419.
 ACCESSION AX456995
 VERSION AX456995.1 GI:21715786
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 1
 AUTHORS Fisher, P.B., Kang, D.C. and Su, Z. Z.
 TITLE Progression suppressed gene 13 (psgen 13) and uses thereof
 JOURNAL Patent: WO 0216419-A 6 28-FEB-2002;
 The Trustees of Columbia University in the City of New York (US)
 FEATURES
 Location/Qualifiers
 1. .835

/organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 246 a 160 c 176 g 253 t
 ORIGIN

Query Match 43.0%; Score 335.2; DB 6; Length 835;
 Best Local Similarity 69.1%; Pred. No. 5,6e-78;
 Matches 539; Conservative 0; Mismatches 218; Indels 23; Gaps 5;

QY 16 GGTCCCTCCCTCTCCACTGACGCTTCTCTTAGCCGACCACTTCTCTTCTTGCT 75
 DB 43 CTTGCTCTCTCTCCACAGCGCTTCTCTTCTTCCGACGCACTTCTCTTCTGCGC 102

QY 76 TGTTCCTCCTTAGGGCGGGAAGCTGAGTGCAGGGTTCAGACCCGCGGAGCAGCTC 135
 DB 103 TCACCTCCCTCCAGTGCAGTGAAGAGTAAACCGGTCAGACCCGCGGAGCTTCT 162

QY 136 TTCAGTGAAGAGGAAGCAATCGGAGGCTGAGCAATGAACCTGAGGATGAGTTAACT 195
 DB 163 CCGCGGGAAGGAACCGCCGACAGAGGAGCAATGATGATGATGATGATGATGAT 222

QY 196 CTTGAGGCGGGAAGGAAGTGTGATGATGATGATGATGATGATGATGATGATGAT 255
 DB 223 CTTTAAAGCTGCAAAACGAAGAGATGTAACATATCCAGGAGAGCTGCTTCTGCAAGG 282

QY 256 GTTTGGGCTCCTCTTCCAGAGCAGACATGTCCTTCTTGAAGAGCTTCTTCTGAGGAG 315
 DB 283 ATTTGGGCTCCTCTTCCGATGATGATGATGATGATGATGATGATGATGATGAT 342

QY 316 TCTGAGGCGGCAAAACGAAGAGATGTTACGTCACGAGGAGAGCTGCTTTTGAAGG 375
 DB 343 TCTTAAAGCTGCAAAACGAAGAGATGTAACATATCCAGGAGAGCTGCTTCTGCAAGG 402

QY 376 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 435
 DB 403 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 462

QY 436 GTA-----TCTGTAACTGGAATAATGATGATGATGATGATGATGATGATGAT 477
 DB 463 GTACTGCCATTTTCTGTAAGAGAGTAACTGGAAT-ATAAGTGAAGAGCAAACTTGA 521

QY 478 AAGTTTCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 535
 DB 522 ACATACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 581

QY 536 CTGTTTTTATACCTTGGAGCAAAACATTAATGATGATGATGATGATGATGATGATGAT 595
 DB 582 CTGCTCTTTTATATCTTGAAGAGAAATCTATGATGATGATGATGATGATGATGATGAT 641

QY 596 TTTTCTTGAAGAGTAACTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 655
 DB 642 ATTTTCTCAGGAATCTGGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 701

QY 656 AAGCTTTTCTTAAACCACTTCTAGTCT--CTGCCACACTTGACACTCCGTCGAAAG 713
 DB 702 GGAATGTTGTTTATATCTTGAAGAGAAATCTATGATGATGATGATGATGATGATGATGAT 761

QY 714 TGAGAAGGAAGTAAAGCAAACTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 773
 DB 762 CAACAAGCAAACTGAAGCAAACTCCTATGAGAAATATATGATGATGATGATGATGATGATGAT 821

RESULT 11
 AC117670 65563 bp DNA linear HTG 10-APR-2002
 LOCUS Mus musculus clone RP23-383P11, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION
 ACCESSION AC117670
 VERSION AC117670.1 GI:20128437
 KEYWORDS HTG; HTGS_PHASED.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 65563)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Mus musculus, clone RP23-383P11
Unpublished
2 (bases 1 to 65563)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23864
Center clone name: 383_P_11

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 693 792: contig of 692 bp in length
* 793 1483: contig of 691 bp in length
* 1484 1583: gap of 100 bp
* 1584 2281: contig of 698 bp in length
* 2282 2381: gap of 100 bp
* 2382 3087: contig of 706 bp in length
* 3088 3187: gap of 100 bp
* 3188 3888: contig of 701 bp in length
* 3889 3988: gap of 100 bp
* 3989 4684: contig of 696 bp in length
* 4685 4784: gap of 100 bp
* 4785 5485: contig of 701 bp in length
* 5486 5585: gap of 100 bp
* 5586 6273: contig of 688 bp in length
* 6274 6373: gap of 100 bp
* 6374 7062: contig of 689 bp in length
* 7063 7162: gap of 100 bp
* 7163 7854: contig of 692 bp in length
* 7855 7954: gap of 100 bp
* 7955 8657: contig of 703 bp in length
* 8658 8757: gap of 100 bp
* 8758 9458: contig of 701 bp in length
* 9459 9558: gap of 100 bp
* 9559 10261: contig of 703 bp in length
* 10262 10361: gap of 100 bp
* 10362 11066: contig of 705 bp in length
* 11067 11166: gap of 100 bp
* 11167 11861: contig of 695 bp in length
* 11862 11961: gap of 100 bp
* 11962 12646: contig of 685 bp in length
* 12647 12746: gap of 100 bp
* 12747 13419: contig of 673 bp in length
* 13420 13519: gap of 100 bp
* 13520 14202: contig of 683 bp in length
* 14203 14302: gap of 100 bp
* 14303 14993: contig of 691 bp in length
* 14994 15093: gap of 100 bp
* 15094 15784: contig of 691 bp in length
* 15785 15884: gap of 100 bp
* 15885 16545: contig of 661 bp in length
* 16546 16645: gap of 100 bp
* 16646 17334: contig of 689 bp in length
* 17335 17434: gap of 100 bp
* 17435 18131: contig of 697 bp in length
* 18132 18231: gap of 100 bp
* 18232 18930: contig of 699 bp in length
* 18931 19030: gap of 100 bp
* 19031 19720: contig of 690 bp in length
* 19721 19820: gap of 100 bp
* 19821 20493: contig of 673 bp in length
* 20494 20593: gap of 100 bp
* 20594 21282: contig of 689 bp in length
* 21283 21382: gap of 100 bp
* 21383 22071: contig of 689 bp in length
* 22072 22171: gap of 100 bp
* 22172 22850: contig of 679 bp in length
* 22851 22950: gap of 100 bp
* 22951 23643: contig of 693 bp in length
* 23644 23743: gap of 100 bp
* 23744 24436: contig of 693 bp in length
* 24437 24536: gap of 100 bp
* 24537 25239: contig of 703 bp in length
* 25240 25339: gap of 100 bp
* 25340 26037: contig of 698 bp in length
* 26038 26137: gap of 100 bp
* 26138 26841: contig of 704 bp in length
* 26842 26941: gap of 100 bp
* 26942 27642: contig of 701 bp in length
* 27643 27742: gap of 100 bp
* 27743 28445: contig of 703 bp in length
* 28446 28545: gap of 100 bp
* 28546 29239: contig of 694 bp in length
* 29240 29339: gap of 100 bp
* 29340 30019: contig of 680 bp in length
* 30020 30119: gap of 100 bp
* 30120 30801: contig of 682 bp in length
* 30802 30901: gap of 100 bp
* 30902 31589: contig of 688 bp in length
* 31590 31689: gap of 100 bp
* 31690 32390: contig of 701 bp in length
* 32391 32490: gap of 100 bp
* 32491 33163: contig of 673 bp in length
* 33164 33263: gap of 100 bp
* 33264 33957: contig of 694 bp in length
* 33958 34057: gap of 100 bp
* 34058 34749: contig of 692 bp in length
* 34750 34849: gap of 100 bp
* 34850 35523: contig of 674 bp in length
* 35524 35623: gap of 100 bp
* 35624 36312: contig of 689 bp in length
* 36313 36412: gap of 100 bp
* 36413 37103: contig of 691 bp in length

TITLE
JOURNAL
COMMENT

[illegible]

QY 459 RAGTTAAAGGACACAATCACTGAATTCCCTTATGTATTTTTATATAGAACCCTTGTAACAAAAA 516
||||| |||| | |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 27541 RAGTTAAAAGACTACTATGTGTAATTTCCTTATGTATTTTTATATAAGAACCTTTTATTATANGA 27598

RESULT 12 AC107839 170882 bp DNA linear HTG 24-AUG-2002
LOCUS Mus musculus clone RP23-284K1, WORKING DRAFT SEQUENCE, 5 ordered pieces.
DEFINITION AC107839

ACCESSION AC107839.4 GI:22474806
VERSION HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS house mouse.
SOURCE Mus musculus, clone RP23-284K1
ORGANISM Unpublished

REFERENCE 2 (bases 1 to 170882)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choquel,Y., Colangelo,M., Collins,S., Collumore,A., Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Farfara,Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glende,S., Gord,S., Guyette,M., Graham,L., Grand-Pierre,N., Grate,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Latocque,K., Lamazzares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., MacLean,C., MacDonald,P., Major,J.J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Miengna,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Petheron,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schubback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stoianovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Vieli,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody.M.

TITLE Direct Submission
JOURNAL Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 170882)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choemel,Y., Collamore,A., Cooke,A., Cooke,P., Desarellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fafra,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine.R., Lindblad-Toh,K., Liu.G., MacLean.C., MacDonald.P., Major,J.J., Matthews.C., McCarthy.M., Meldrim,J., Meneus.L., Mihova.T., Mienga.V., Murphy.T., Naylor.J., Nguyen.C., Nicol.R., Norbu.C., Norman.C.H., O'connor.T., O'donnell.P., O'Neill.D., Oliver.J., Peterson.K., Phunkhang.P., Pierre.N., Raymond.C., Retta.R., Rise.C., Rogov.P., Roman.J., Roy.A., Schauer.S., Schubback.R., Seaman.S., Severy.P., Smith.C., Spencer.B., Stange-Thomann.N., Stoianovic.N., Talamas,J., Tesfaye.S., Theodore.J., Toptom.K., Travers.M., Vassiliev.H., Vieli.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Young.G., Zainoun.J., Zemdek.L., Zimmer.A. and Zody.M.

TITLE Direct Submission
JOURNAL Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Aug 24, 2002 this sequence version replaced gi:20163124.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTHR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L20600

Center clone name: 284_K.1

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 169045 bases at least Q40

Consensus quality: 169800 bases at least Q30

Consensus quality: 170139 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 170482; sum-of-contigs

Quality coverage: 13.6 in Q20 bases; agarose-fp

Quality coverage: 13.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 107625: contig of 107625 bp in length

* 107626 107725: gap of 100 bp

* 107726 111441: contig of 3716 bp in length

* 111442 111541: gap of 100 bp

* 111542 120063: contig of 8522 bp in length

* 120064 120163: gap of 100 bp

* 120164 161248: contig of 41085 bp in length

* 161249 161348: gap of 100 bp

* 161349 170882: contig of 9534 bp in length.

FEATURES
source

1..170882

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="RP23-284K1"

/clone.lib="RPC1-23 Female Mouse BAC"

1..107625

/note="assembly_fragment"

clone_end:SP6

vector_side:left

107726..111441

/note="assembly_fragment"

111542..120063

/note="assembly_fragment"

120164..161248

/note="assembly_fragment"

161349..170882

/note="assembly_fragment"

clone_end:T7

vector_side:right

a 37012 c 37190 g 44751 t 400 others

BASE COUNT

ORIGIN

Query Match 42.8%; Score 334.2; DB 2; Length 170882;
Best Local Similarity 89.4%; Pred No. 1.6e-77;
Matches: 371; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

QY 99 CTGAGTGCAGGGTTCAGACCCAGCGGCGAGCAGCTCTTCAGTGAAGAGGAGCAATCG 158

|||||
DB 56621 CAGGGTACAGGGCTCAGACCCAGCAGC-AGCAGCTCTCCGTCGAGAGGAGCAATCG 56563

QY 159 GAGGGTCAGCAATGACGTGGAGCATGAGTTAACTCTCGTGGGAGGAATTCATCGTC 218

|||||
DB 56562 GAGAGTCAGCAATGATGTGGAACATGAAGTTAACTCTCGTGGGAGGAATTAATCGCC 56503

QY 219 TGGGTTCCAAAATGCCGATGGGAAACTGAGTGTGAAGTTGGGGTCTCTTCCAAAGACG 278
|||||
DB 56502 TGGGTTCCAAAATGCTGATGGAAATTAAGTCTGAAGTTGGGGTCTCTTCCAGAAATG 56443
|||||
QY 279 ACAGATGTCCTATCTCTTTGAAGCGTTGGTGGAACTCTGAAGCGGCAAAACGAGGA 338
|||||
DB 56442 ACAGATGTCCTATCTCTTTGAAGCAATGGTAGAACTCTGAAGCTGCAAAACGAGGA 56383
|||||
QY 339 AGATTCTTACGTACGAGGAGAGCTGCTTTTGCAGAGTGTTCATGATGATGTTGACATTTG 398
|||||
DB 56382 AGATTCTTACATACACAGGGAACTACTTTTGAAGTGTTCATGATGATGTTGACATTTG 56323
|||||
QY 399 TATTGTCGAAGATTAATGTGTTTCAGATCTGGGGTATCTGTTAACTGGAATAATT 458
|||||
DB 56322 TATTGTCGAAGATTAATGTGTTTCAGATCTGGGGTATCTGTTAACTGGAATAATT 56263
|||||
QY 459 AGTTAAAGGACAAACATGAAGTTCCCTATGATATTTTATAGACCTTTGTAACA 513
|||||
DB 56262 AAGTTAAAGACTAATGTGAATTCCTATGATATTTTATAGAACTTTTATATA 56208
|||||

RESULT 13
AF116682 standard; RNA; HDN; 876 BP.
XX AF116682;
XX AF116682.1
SV AF116682.1
XX
DT 24-MAY-2000 (Rel. 63, Created)
DT 09-MAY-2001 (Rel. 67, Last updated, Version 2)
XX
DE Homo sapiens PRO2013 mRNA, complete cds.
XX
KW
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN [1]
RP Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,
RA Liu M., He F.;
RT "Functional prediction of the coding sequences of 121 new genes deduced by
analysis of cDNA clones from human fetal liver";
RL Unpublished.
XX
RN [2]
RP Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,
RA Liu M., He F.;
RT
RL Submitted (24-DEC-1998) to the EMBL/GenBank/DBJ databases.
RL Department of Experimental Hematology, Institute of Radiation Medicine,
Beijing Taiping Road 27, Beijing 100850, P. R. China
XX
DR SPTREMBL; Q9P1F3; Q9P1F3.
XX
FH Key Location/Qualifiers
FT source
FT 1..876
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="FLB7547"
FT /tissue_type="liver"
FT /dev_stage="fetal"
FT 136..381
FT CDS
FT /codon_start=1
FT /db_xref="SPTREMBL:Q9P1F3"
FT /note="Predicted protein of HQ2013"
FT /product="PRO2013"


```

|||||
237 GAAGGAAGATGTACATATCAGGAGAGCTCTTCTGCAAGGTGTTTCATGATGATGTTG 296
Qy |||||
393 ACATTGTTATGTCGAAGATTAATGTTTGCAGATCTGGGGTA----- 438
Db |||||
297 ACATTATATTTACTGCAAGATTAATGTTTACATATCTTTATGTCACGCCATTTTGTCT 356
Qy |||||
439 -TCGGTAACTGGATTAATTAAGTAAAGCAACAACAT---GAAGTCTCTTATGATTTT 494
Db |||||
357 TCTGTTAACTGGAAT-ATAAGTGAAGAACAAACATTTTGAACATACTTAATGATTTT 415
Qy |||||
495 TTATAGACCTTTGTAACAAAGGGA--CTGTTGAGAGTCCTGTTTATACCTTGG 552
Db |||||
416 TTATAGAACTTTGTAAAGGAGAGATTCATGTTTGAAGTCTGTCTTTTATATC 475
Qy |||||
553 AGCAAAACATACATGTAATAAATAAACAACCTGTTATTTTCTTCTTAAAGAGCT 612
Db |||||
476 TTGAAGAAATATGATGATGCTATATAAATAATCCTATATTTTCTCAGGAATCT 535
Qy |||||
613 AATCGGAGAGCTAGGCAATTAATGTTTTCAGAGGTGCGAAAAGCTTTTCTTTCTTA 672
Db |||||
536 GGTAGGAATTTGCAGGCAATGAGATTTTTCGGGGCAGGATGGGAATGTTTGTTCATA 595
Qy |||||
673 AACCATCTTAGTCT--CTGCCACACTTGACACTCCGTCGAAGTGAAGCAACATAAG 730
Db |||||
596 AATATATAGACATTTCTATAGATATTTGACATCTGCGAAGCAACAACTGAAG 655
Qy |||||
731 ACCAACTCGGTGGAATAATATGTTTATGTAATAAAAAA 773
Db |||||
656 ACCAACTCTATGAGAAATATATGATGTTATGTAATAAAGA 698

RESULT 15
AL590308
LOCUS Human DNA sequence from clone RP11-501K14 on chromosome 6, complete
sequence.
ACCESSION AL590308
VERSION AL590308.8 GI:14596398
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 100296)
Blakey, S.
Direct Submission
Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Jul 4, 2001 this sequence version replaced gi:14252463.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6

```

RP11-501K14 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-501K14. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-501K14 is at 1 in this sequence.

The true left end of clone RP1-225E12 is at 98297 in this sequence.

The true right end of clone RP11-94L3 is at 97304 in this sequence.

FEATURES

repeat_region	source
1..100296	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="6"
	/clone="RP11-501K14"
	/clone_lib="RPCI-11.2"
2..99	
	/note="MIR repeat: matches 103..194 of consensus"
118..199	
	/note="MER47 repeat: matches 2242..2323 of consensus"
190..265	
	/note="MER47 repeat: matches 2242..2317 of consensus"
268..346	
	/note="MER47 repeat: matches 1..79 of consensus"
502..671	
	/note="MIR repeat: matches 7..185 of consensus"
1420..1737	
	/note="AluJb repeat: matches 2..310 of consensus"
2010..2033	
	/note="12 copies 2 mer tt 100% conserved"
2650..2961	
	/note="AluX repeat: matches 3..306 of consensus"
5767..5895	
	/note="MIR repeat: matches 21..166 of consensus"
6062..6159	
	/note="L2 repeat: matches 2587..2709 of consensus"
6165..6268	
	/note="L2 repeat: matches 2607..2709 of consensus"
6916..7157	
	/note="AluJb repeat: matches 1..295 of consensus"
7301..7442	
	/note="L2 repeat: matches 1768..1916 of consensus"
7746..7999	
	/note="AluX repeat: matches 45..298 of consensus"
8013..8152	
	/note="MIR repeat: matches 14..151 of consensus"
8899..8959	
	/note="MIR repeat: matches 194..256 of consensus"
9209..9497	
	/note="L1MB8 repeat: matches 5863..6175 of consensus"
9503..9698	
	/note="AluJb repeat: matches 90..281 of consensus"
9703..10428	
	/note="L1MB8 repeat: matches 5126..5867 of consensus"
10429..10759	
	/note="AluJb repeat: matches 1..310 of consensus"
10760..10773	
	/note="L1MB8 repeat: matches 5114..5126 of consensus"
10897..11050	
	/note="L2 repeat: matches 821..962 of consensus"
11051..11360	
	/note="AluSg repeat: matches 1..310 of consensus"
11361..12691	
	/note="L2 repeat: matches 962..2419 of consensus"
12850..13009	
	/note="AluJb repeat: matches 129..279 of consensus"
15295..15341	
	/note="MIR repeat: matches 97..142 of consensus"
15342..15632	
	/note="AluSg repeat: matches 21..308 of consensus"
15633..15735	
	/note="MIR repeat: matches 142..250 of consensus"

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (Bases 1 to 551)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.
, Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other ESTs: id05d03.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brownjefas.harvard.edu)
MG1:1948459 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 429.
Location/Qualifiers
1. 551
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5662133"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
N1-MMS1"
/sex="Both for embryonic & newborn, male for adult and
adult islet"
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="DH10B"
/note="Vector: SPORT1; Site_1: Not I; Site_2: Sal I; Five
libraries representing E10.5/12.5 pancreatic bud, E16.5
pancreas, newborn pancreas, adult pancreas, and adult
islets of Langerhans were separately constructed using
SuperScript plasmid library kit (Life Technologies). cDNA
was made by oligo-dT priming and size-selected by column
fractionation. Libraries were amplified once on solid
support and plasmid DNA from each library was prepared
and mixed in equal amounts. The mixed library DNA was
normalized by method #4 from Bonaldo, Lennon, and Soares
1996 Genome Research 6:791-806; 0.5 microgram
single-stranded mixed library plasmid DNA was mixed with
5 micrograms PCR product representing mixed library
inserts and hybridized to an Ecot of 6. Single-stranded
(unhybridized) plasmids were isolated by hydroxyapatite
chromatography and used to make this library."
BASE COUNT 163 a 100 c 133 g 155 t
ORIGIN
Query Match 58.6%; Score 456.8; DB 13; Length 551;
Best Local Similarity 92.08; Pred. No. 2e-104;
Matches 504; Conservative 0; Mismatches 42; Indels 2; Gaps 2;
QY 52 CCCGAACCACTTCCTTCTTCTGCTTCTCTCCCTAGGCGCGGAAGCTCAGTGCAGGTT 111
|||||
Db 1 CCCGGGCCACTTCCTTCTTCTGCTTCTCTCCCGAGGCGCGGAAGCTCAGTGCAGGTT 111
|||||
QY 112 TCAGACCCAGCGGAGCAGCTCTTTTCAGTGAAGAAGAAAGCAATCGAGGGTCAGCAAT 171
|||||
Db 61 TCAGACCCAGCGCAGC-AGCAGCTCTTCCGTTGAAGAAGAAAGCAATCGAGAGTCAGCAAT 119
|||||
QY 172 GAAGTGTGGACATCAGGTTAACTCTCGTGTGAGAGAAATTCATCGTCTGGGTTCCAAAA 231
|||||

[illegible]

Db	309	CAAGGTGTCATGATGATGTCATGTCGCAAGATTAATGTTGGTTGCATGGC	368
Qy	431	TGGGGGTATCTGGTAAACTGGAATAATTAAGTTAAAGGACAAACATGAAGTTCCTTATGT	490
Db	369	TTCGTGTCATCTGATAAACTTGGATAAACAAGTTAAAGACTAACGTGAATTTCCCTTATGT	428
Qy	491	ATTTTATTAGACCTTTCTGAACRAAAGGGGACTTGTGAGAAGTCCCTGTTTTATACCTT	550
Db	429	ATTTTATTAGAACCTTTGTAACAAAGGGGGCTTGTGAGAAGTCCCTGTTTTATACCTT	488
Qy	551	GGAGCAAAACATTACAATG	569
Db	489	GAGCAAAACATTACAATG	507

RESULT	12
LOCUS	BQ445291
DEFINITION	UI-M-ERO-bxn-b-05-0-UI.r1 NIH_BMAP_ERO Mus musculus cDNA clone
ACCESSION	BQ445291
VERSION	BQ445291.1
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. James Lin, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ LLNL at: http://image.llnl.gov This clone was contributed by the Brain Molecular Anatomy Project (BMAP) Seq primer: pyx-5.

[illegible]

RESULT 13
 BE689324
 LOCUS
 DEFINITION
 BE689324 504 bp mRNA linear EST 11-SEP-2000
 uw52c09.y1 Soares_mammary_gland_MLMLG Mus musculus CDNA clone
 IMAGE:3465616 5' similar to TR:O81881 O81881 HYPOTHETICAL 18.3 KD
 PROTEIN, ;, mRNA sequence.
 ACCESSION
 BE689324
 VERSION
 BE689324.1 GI:10076948
 KEYWORDS
 EST.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE
 1 (bases 1 to 504)
 AUTHORS
 NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL
 Unpublished (1997)
 COMMENT
 Other_ESTs: uw52c09.x1
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mai.nih.gov
 This clone is available royalty-free through LML ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:1384976
 Seq primer: -40RP from Gibco
 High quality sequence stop: 471.
 Location/Qualifiers
 1..504
 /organism="Mus musculus"
 FEATURES
 source

	BASE COUNT	ORIGIN
1	0.000	0.000
2	0.000	0.000
3	0.000	0.000
4	0.000	0.000
5	0.000	0.000
6	0.000	0.000
7	0.000	0.000
8	0.000	0.000
9	0.000	0.000
10	0.000	0.000
11	0.000	0.000
12	0.000	0.000
13	0.000	0.000
14	0.000	0.000
15	0.000	0.000
16	0.000	0.000
17	0.000	0.000
18	0.000	0.000
19	0.000	0.000
20	0.000	0.000
21	0.000	0.000
22	0.000	0.000
23	0.000	0.000
24	0.000	0.000
25	0.000	0.000
26	0.000	0.000
27	0.000	0.000
28	0.000	0.000
29	0.000	0.000
30	0.000	0.000
31	0.000	0.000
32	0.000	0.000
33	0.000	0.000
34	0.000	0.000
35	0.000	0.000
36	0.000	0.000
37	0.000	0.000
38	0.000	0.000
39	0.000	0.000
40	0.000	0.000
41	0.000	0.000
42	0.000	0.000
43	0.000	0.000
44	0.000	0.000
45	0.000	0.000
46	0.000	0.000
47	0.000	0.000
48	0.000	0.000
49	0.000	0.000
50	0.000	0.000
51	0.000	0.000
52	0.000	0.000
53	0.000	0.000
54	0.000	0.000
55	0.000	0.000
56	0.000	0.000
57	0.000	0.000
58	0.000	0.000
59	0.000	0.000
60	0.000	0.000
61	0.000	0.000
62	0.000	0.000
63	0.000	0.000
64	0.000	0.000
65	0.000	0.000
66	0.000	0.000
67	0.000	0.000
68	0.000	0.000
69	0.000	0.000
70	0.000	0.000
71	0.000	0.000
72	0.000	0.000
73	0.000	0.000
74	0.000	0.000
75	0.000	0.000
76	0.000	0.000
77	0.000	0.000
78	0.000	0.000
79	0.000	0.000
80	0.000	0.000
81	0.000	0.000
82	0.000	0.000
83	0.000	0.000
84	0.000	0.000
85	0.000	0.000
86	0.000	0.000
87	0.000	0.000
88	0.000	0.000
89	0.000	0.000
90	0.000	0.000
91	0.000	0.000
92	0.000	0.000
93	0.000	0.000
94	0.000	0.000
95	0.000	0.000
96	0.000	0.000
97	0.000	0.000
98	0.000	0.000
99	0.000	0.000
100	0.000	0.000

Query Match

53.38; Score 415.6; DB 14; Length 547;

/sex="male"
 /tissue_type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)
 primer [5].
 TGTTCACCAATCTGAAGTGGGCGCGCGCAATGTTTGTGTGTGTGTGTGTGTGT
 T 3'; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT7T3 vector.
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
 constructed and normalized by Bento Soares and M.Fatima
 Bonaldo."

BASE COUNT 185 a 96 c 143 g 167 t

Query Match 54.5%; Score 425; DB 9; Length 591;

Best Local Similarity 88.7%; Pred. No. 2.le-96;

Matches 496; Conservative 0; Mismatches 55; Indels 8; Gaps 3;

QY 103 GTGAGGTTTCAGACCCAGCGCGAGCAGCTCTTCAGTGAAGAGGAGCAATCGGAGG 162
 Db 7 GTACAGGTTTCAGACCCAGCGAGC-AGCAGCTCTTCGTGAAGAGGAGCAATCGGAG 65
 QY 163 CTCAGCAATGAACGTGGAGCATGAGGTTAACTCTCTGTGGAGGAAATTCATCGTCTGGG 222
 Db 66 CTCAGCAATGAATGGAGCAATGAGTTAACTCTCTGTGGAGGAAATTCATCGTCTGGG 125
 QY 223 TTCAAAATGCCAGTGGAACTGAGTGTGAAGTTTGGGTTCCTTCCAGAGCAGAC 282
 Db 126 TTCAAAATGCTGATGAAAAATTAAGTGTGAAGTTTGGGTTCCTTCCAGAGTACAG 185
 QY 283 ATGTGCCAATCTTTGAAGCGTTGGTGGAGTCTGAAGCCGCAAAACGAGGAAGAT 342
 Db 186 ATGTGCCAATCTTTGAAGCGTTGGTGAAGTCTGAAGTCTGAAGTCTGAAGTCTGAAG 245
 QY 343 TGTTCAGTACGAGGAGAGTCTTTTCAAGGTTTTCATGATGATGTTGACATTTGATT 402
 Db 246 TGTTCATACGAGGAGGAACTATTTTCAAGGTTTTCATGATGATGTTGACATTTGATT 305
 QY 403 GCTCCAGATTAATGTGTTTGCAGATCTGGGATCTGTTAACTGGAATAATTAAGT 462
 Db 306 GCTCCAGATTAATGTGTTTGCATGTTGTTGATGTTGATGTTGATGTTGATGTTGAT 365
 QY 463 TAAAGGACAAACATGAAGTCTCTTATGATTTTATAGACCTTTCTGAACAAAGGGAC 522
 Db 366 TAAAGACTACCTGAATTTCTTATGATTTTATAGACCTTTCTGAACAAAGGGGAC 425
 QY 523 TGTGTGAAGTCTCTGTTTATACCTTGGAGCAAAACATTACATGTAATAAATAACAA 582
 Db 426 TGTGTGAAGTCTCTGTTTATACCTTGAAGCAAAACATTACATGTAATAAATAACAA 485
 QY 583 AACCTGTTATTTTTCCTTAAGAAGTAAATCGGGAGCTAGGCAATAAATGTTTT 642
 Db 486 ACC-----TATTATTTTCTTAAGAAGTAAATGGGAATGTAGTAAAT-GAACATTTT 538
 QY 643 CAGAGTTCGAAAAGCTT 661
 Db 539 TGGAGGTGTGAAGGCTT 557

RESULT 11

AA008262

LOCUS

DEFINITION

mg/7901.r1 Soares mouse embryo NM013.5 14.5 Mus musculus cDNA

clone IMAGE:439056 5', mRNA sequence.

ACCESSION

AA008262

VERSION

AA008262.1 GI:1464233

KEYWORDS

EST.

SOURCE

house mouse.

Mus musculus

REFERENCE AUTHORS

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 507)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geiseli,S., Kucaha,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE JOURNAL COMMENT

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:264392

Seq primer: EYPrimer

High quality sequence stop: 359.

FEATURES source

Location/Qualifiers

1..507

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_image="IMAGE:439056"

/clone_lib="Soares mouse embryo NM013.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5].
 TGTTCACCAATCTGAAGTGGGCGCGCGCAATGTTTGTGTGTGTGTGTGTGTGT
 T 3'; on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT7T3 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Fatima Bonaldo."

BASE COUNT 148 a 87 c 133 g 138 t 1 others

ORIGIN

Query Match

Best Local Similarity 91.4%; Score 420.4; DB 9; Length 507;

Matches 456; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 71 CTGCTTGTCTCCCTAGGGCGGGAAGTGTAGTCAGGGTTCAGACCCAGCGGAGC 130

Db 10 CNGAGGTTTCTCCCGAGGCGCGGAACCTGTAGTACAGGGTTCAGACCCAGCAGC -AGC 68

QY 131 AGCTCTTCAGTGAAGAGGAGCAATCGAGGCTCAGCAATGACGTGGAGCATGAGTT 190

Db 69 AGCTCTTCGTGAAGAGGAGCAATCGAGGCTCAGCAATGACGTGGAGCATGAGTT 128

QY 191 AACCTCTGTGGAGGAAATTCATCTCTGGGTTCACAAATCCGATGGAACATGAGT 250

Db 129 AACCTCTGTGGAGGAAATTCATCTCTGGGTTCACAAATCCGATGGAACATGAGT 188

QY 251 GTGAAGTTTGGGTCTCTTCCAGAGCAGAGATGTGCCAATCTCTTTGAAGGTTGGTG 310

Db 189 GTGAAGTTTGGGTCTCTTCCAGAGTACAGATGTGCCAATCTCTTTGAAGGTTGGTA 248

QY 311 GGAACCTCTGAAGCCCAAAACAGAGAGATTGTTTACGTACGAGAGAGCTGCTTTTG 370

Db 249 GGAACCTCTGAAGCTCAAAACAGAGAGATTGTTTACATACGAGGGAACACTTTTG 308

QY 371 CAAGGTGTTTCATCATGATGTTGACATTTGATTTCTCTCAAGATTAAATGTTTCCAGATC 430

```

/db_xref="taxon:10090"
/clone="IMAGE:3465616"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 156 a 80 c 123 g 144 t 1 others
ORIGIN
Query Match 53.2%; Score 414.6; DB 10; Length 504;
Best Local Similarity 91.6%; Pred. No. 8.6e-94;
Matches 460; Conservative 0; Mismatches 40; Indels 2; Gaps 2;
QY 98 GCTGAGTGCAGGTTTCAGACCCGCGGAGCAGCTCTTCAGTCAAGAGGAAGCAATC 157
Db 1 GCCGAGTACAGGTTTCAGACCCGCGGAGCAGCTCTTCAGTCAAGAGGAAGCAATC 59
QY 158 GGAGGGTTCAGCAATGAGCTGGAGCATGAGGTTAACCTCCCTGGTGGAGGAATTCATCGT 217
Db 60 GGAGAGTCAGCAATGAGCTGGAGCATGAGGTTAACCTCCCTGGTGGAGGAATTCATCGC 119
QY 218 CTGGGTTCCAAAATGCGGATGGGAACCTGAGTGTGAAGTTTGGGGTCTCTTCCAAAGC 277
Db 120 CTGGGTTCCAAAATGCGGATGGGAATTAAGTGTGAAGTTTGGGGTCTCTTCCAGGAT 179
QY 278 GACAGATGTGCCATCTCTTTGAGCGTTGGTGGAACTCTGAAAGCGCGCAAGG 337
Db 180 GACAGATGTGCCATCTCTTTGAGCGTTGGTGGAACTCTGAAAGCGCGCAAGG 239
QY 338 AAGATTGTACGTACGAGGAGAGCTGCTTTTGCAGGTGTTTCATGATGATGATGATGATG 397
Db 240 AAGATTGTACGTACGAGGAGAGCTGCTTTTGCAGGTGTTTCATGATGATGATGATGATG 299
QY 398 GTATTGTGCAAGATTAATGTGGTTTGCAGATCTGGGGGTATCTCGTAACTGGAATAAT 457
Db 300 GTATTGTGCAAGATTAATGTGGTTTGCAGATCTGGGGGTATCTCGTAACTGGAATAAC 359
QY 458 TAAGTTAAAGACAAACATGAGTTCCTTATCTATTTTATAGACTTGTGAACAAAG 517
Db 360 TAAGTTAAAGACAAACATGAGTTCCTTATCTATTTTATAGACTTGTGAACAAAG 419
QY 518 GGGACTTGTGAGAGTCTCTGTTTATACCTTGGAGCAAAACATTAACAATGTAATAATA 577
Db 420 GGGCTTGTGAGAGTCTCTGTTTATACCTTGGAGCAAAACATTAACAATGTAATAATA 478
QY 578 AAAAAACCTGTTATTTTTTTT 599
Db 479 AAAAAACCTGTTATTTTTTTT 500

```

```

RESULT 14
BF468668 501 bp mRNA linear EST 04-DEC-2000
LOCUS UI-M-BH3-atu-a-11-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
DEFINITION UI-M-BH3-atu-a-11-0-UI 5', mRNA sequence.
ACCESSION BF468668
VERSION BF468668.1 GI:11537851
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 501)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
TITLE

```

```

discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.
Location/Qualifiers
1. .501
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="UI-M-BH3-atu-a-11-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_host="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)."
BASE COUNT 124 a 115 c 123 g 139 t
ORIGIN
Query Match 53.0%; Score 413.6; DB 12; Length 501;
Best Local Similarity 91.8%; Pred. No. 1.5e-93;
Matches 448; Conservative 0; Mismatches 39; Indels 1; Gaps 1;
QY 9 CTCTCTCTGTCCTCCCTTCTCCACTGCAGCCTTCTCTTAGCCGCAACACTTCCTTC 68
Db 15 CTCTCTCTGTCCTCCCTCCCTCCACTGCAGCCTTCTCTTAGCCGCAACACTTCCTTC 74
QY 69 TTCTGCTTGTCTCTCCCTAGGCGCGGAAGCTGAGTGCAGGTTTCAGACCCACCGCGGA 128
Db 75 TTCTGCTTGTCTCTCCCTAGGCGCGGAAGCTGAGTGCAGGTTTCAGACCCACCGCAGC-A 133
QY 129 GCAGCTCTTCAGTGAAGAGGAAGCAATCGAGGCTGAGCAATGAACGTGAGCATGAGG 188
Db 134 GCAGCTCTTCCTGAGAGAGGAAGCAATCGAGGCTGAGCAATGAATGTGGAACATGAAG 193
QY 189 TTAACCTCTGTTGGAGGAATTCATCGTGGGTTCCAAAATGCGGAGTGGGAACTGA 248

```

Db 194 TTAACCTCCTGCTGGAGGAAATTCATCGCTGGGTTCCAGAAATGCTGATGGAATAATTA 253
 QY 249 GTGTGAAGTTTGGGGTCCCTCTTCCAGAGACGACAGATGTGCCAATCTCTTTTGAAGCGTTGG 308
 Db 254 GTGTGAAGTTTGGGGTCCCTCTTCCAGAGATGACAGATGTGCCAATCTCTTTTGAAGCGTTGG 313
 QY 309 TGGGAACCTCTGAAGCGGCAAAACGAGGAAGATTTGCTAGTCGCGAGGAGAGCTGCTTT 368
 Db 314 TAGGAACCTCTGAAGCGTCAAAACGAGGAAGATTTGCTAGTCGCGAGGAGAGCTGCTTT 373
 QY 369 TCGAAGCTTTCATGATGATGTGACATGTATGCTGCAAGATTAATGTGGTTTGCAGA 428
 Db 374 TCGAAGCTTTCATGATGATGTGACATGTATGCTGCAAGATTAATGTGGTTTGCATG 433
 QY 429 TCTGGGGTATCTGTGTAACCTGGAATAATTAAGTTTAAAGCAAAACGAGGAAGTTTCCCTTAT 488
 Db 434 GCTTGTGTGATGATAACTGGAATAACTTAAGTTTAAAGCAAAACGAGGAAGTTTCCCTTAT 493
 QY 489 GTATTTTT 496
 Db 494 GTATTTTT 501

RESULT 15
 BM122169/c
 LOCUS
 DEFINITION
 L0506D12-3 NIA Mouse Newborn Heart cDNA Library Mus musculus cDNA
 clone L0506D12 3', mRNA sequence.
 ACCESSION
 BM122169
 VERSION
 BM122169.1
 KEYWORDS
 EST.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE
 1 (bases 1 to 580)
 Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.
 and K.O.M.S.H.
 Systematic Analyses of NIA Mouse Newborn Heart cDNA Library
 Unpublished (2001)
 CONTACT
 Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 Plate: L0506 row: D column: 12
 Seq primer: -21M13 Forward
 High quality sequence stop: 580
 POLYA=Yes.

FEATURES
 source
 1..580
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="niaEST:L0506D12-3"
 /db_xref="taxon:10090"
 /clone="L0506D12"
 /clone_lib="NIA Mouse Newborn Heart cDNA Library"
 /tissue_type="Newborn Heart"
 /dev_stage="Newborn"
 /lab_host="DH10B"
 /note="Organ: heart; Vector: pSPORT1 (Invitrogen); Site: 1:
 SalI; Site: 2: NotI; Mouse cDNA project by the Laboratory
 of Genetics, National Institute on Aging (NIA), Intramural
 Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).
 Double-stranded cDNAs were synthesized with an Oligo(dT)
 primer [Invitrogen: 5'-
 PGACTAGTCTAGATCGGAGCGCGCCCTTTTCTTTTCTTTT-3'] from
 24.9 microgram of total RNA, treated with T4 DNA
 polymerase, and purified by ethanol-precipitation. The
 cDNAs were ligated to Lone-linker LL-sal3 (Ref.
 Development 127:1737-1749 (2000) [PMID: 10725249]),
 purified by phenol/chloroform, and separated from free
 linkers by Centricon 100. Then, the cDNAs were digested

with SalI and NotI enzymes, and cloned into SalI and NotI
 site of pSPORT1 plasmid vector. The DH10B E. coli host was
 transformed with ligation mixture by the chemical method.
 The average insert size is about 1.8 kb. The library was
 constructed by Yulan Piao (NIA)."

BASE COUNT 177 a 127 c 82 g 194 t
 ORIGIN
 Query Match 52.18; Score 406.6; DB 13; Length 580;
 Best Local Similarity 85.6%; Pred. No. 8.8e-92;
 Matches 492; Conservative 0; Mismatches 69; Indels 14; Gaps 3;
 QY 205 GGAATTCATCGTCTGGGTTCCAAAATGCCGATGGGAACCTGAGTGTGAAGTTTGGGT 264
 Db 580 GGAATTCATCGCTGGGTTCCAGAAATGCTGATGGAATTAAGTGTGAAGTTTGGGT 521
 QY 265 CCTCTTCAAGACGACAGATGTGCAATCTCTTTGAAGCGTTGGGGAACCTCTGAAAGC 324
 Db 520 CCTCTTCAAGACGACAGATGTGCAATCTCTTTGAAGCGTTGGGGAACCTCTGAAAGC 461
 QY 325 CGCAAAAGGAAGGATTTAGTACGACGAGGAGAGCTCTTTTGCAGGTTTCATGA 384
 Db 460 TGCAAAAGGAAGGATTTAGTACGACGAGGAGAGCTCTTTTGCAGGTTTCATGA 401
 QY 385 TGATCTTGACATTTGCTGCTCAAGATTAATGTGTTTGCAGATCTGGGGGTATCTGCT 444
 Db 400 TGATGTTGACATTTGCTGCTCAAGATTAATGTGTTTGCAGATCTGGGGGTATCTGCT 341
 QY 445 AAATCGGAATTAAGTTTAAAGGACAAACATGAAGTTTCTTATCTATTTTATAGACT 504
 Db 340 AAATCGGAATTAAGTTTAAAGGACAAACATGAAGTTTCTTATCTATTTTATAGACT 281
 QY 505 TTGTAAACAAAGGGGACTTCTTGAGAAGTCTCTGTTTATACCTTGGAGCAAAACATTA 564
 Db 280 TTGTAAACAAAGGGGACTTCTTGAGAAGTCTCTGTTTATACCTTGGAGCAAAACATTA 221
 QY 565 CAATGTAATAAATAACAAACCTGTTATTTTCTTAAAGAGGTAATCGGGAGAGC 624
 Db 220 CAATGTAATAAATAACAAACCTGTTATTTTCTTAAAGAGGTAATCGGGAGAGC 167
 QY 625 TAGCAATAAATGTTTTCAGAGTGGCAAAAGCTTTTGTCTTAAACCAATCTT-- 682
 Db 166 TAGCAATAAATGTTTTCAGAGTGGCAAAAGCTTTTGTCTTAAACCAATCTTAA 107
 QY 683 ---AGTCTCTGCCACACTTGACACTCGTCAAGTGAAGCGGAGTAAAGCAACTG 738
 Db 106 GACAATTTCTACAGGCACTTGACATCTGTCAAGCAAGAAGCAAACTGCAGACCGCTG 47
 QY 739 CGGTGGAATAATATTATGTTTATTAATAAAAAA 773
 Db 46 CCATGGAATAATGTT--TGTTTATGGAATAAAAAA 14

Search completed: January 17, 2003, 02:00:47
 Job time : 2236 secs

```
putative"
/codon_start=1
/protein_id="BAB29100.1"
/db_xref="GI:12851587"
/db_xref="MGD:MGI:1920362"
/translation="NMVEHVNLLVEEIIHRLSGNRADGKLSKFGVLFODDRCANLFE
ALVGLKAAKRKIIVTYAGELLQGVHDDVDIVLLQD"
polyA_signal
746..751
/note="putative"
polyA_site
769
/note="putative"
BASE COUNT 228 a 146 c 174 g 221 t
ORIGIN
Query Match 69.1%; Score 539.2; DB 11; Length 769;
Best Local Similarity 86.6%; Pred. No. 3.6e-125;
Matches 658; Conservative 0; Mismatches 83; Indels 19; Gaps 5;
QY 20 CCTCCCTCTCCACGACGACCTTCTCTTAGCCGCAACCACTCTCTCTCTCTCTGTT 79
DB 10 CCTCCCTCTCCACGACGACCTTCTCTTAGCCGCGGCACTCTCTCTCTCTCTGTT 69
QY 80 CCTCCCTAGCGCGGCGAAGCTGAGTGCAGGTTTCAGACCCACGCGGAGCAGCTCTCA 139
DB 70 CCTCCCGGCGGCGAAGCGAGTACAGGTTTCAGACCCACGCGAGC-AGCAGCTCTCC 128
QY 140 GTGAAGAAGGAAGCAATCGGAGGCTGAGCAATGAAGCTGAGGATGAGTTAACTCTCTG 199
DB 129 GTGAAGAAGGAAGCAATCGGAG-AGCAATGAATGTGAACATGAAGTTAACTCTCTG 184
QY 200 GTGGAGGAATATCATCTGCTGGTGTTCAGAAATGCGGATGCGGAACTGAGTGTGAAGTTT 259
DB 185 GTGGAGGAATATCATCTGCTGGTGTTCAGAAATGCTGAGAAATTAAGTGTGAAGTTT 244
QY 260 GGGGTCTCTTCCAGACGACAGATGTGCAATCTCTTTCAGGCTGTGGGGAACCTCTG 319
DB 245 GGGGTCTCTTCCAGATGNCAGATGTGCAATCTCTTTCAGGCTGTGGGGAACCTCTG 304
QY 320 AAAGCGGCAAAAGGAAGATGTTACGTACGACGAGAGCTGCTTTTTCGCAAGTGTGT 379
DB 305 AAAGCTGCAAAAGGAAGATGTTACATACGCGGGAAGCACTACTTTTTCGCAAGTGTGT 364
QY 380 CATGATGATGTCACATGTTATGCTGCAAGATTAATGCTGTTGAGATCTGGGGGTAT 439
DB 365 CATGATGATGTTGACATGTTATGCTGCAAGATTAATGCTGTTGAGATCTGGGGGTAT 424
QY 440 CTGGTAACTGGAATAATTAAGTTAAAGGCAACATGAAGTTCCTTATCTATTTTATA 499
DB 425 CTGATAACTGGATACTAAGTTAAAGACTACGTGAATTTCCCTTATGATTTTATA 484
QY 500 GACCTTTGTAACAAAGGGGACTGTTGAGAGTCTCTGTTTATACCTTTGGAGCAAAA 559
DB 485 GAACCTTTGTAACAAAGGGGCTGTTGAGAGTCTCTGTTTATACCTTTGAAGCAAAA 544
QY 560 CATTCAATGTAATAATAACAACTGTTATTTTTCCTTAAGAGGTAATCGGG 619
DB 545 CATTCAATGTAATAATAACAACTGTTATTTTTCCTTAAGAGGTAATCGGG 598
QY 620 AGACGTAGGCAATAAATGTTTTCAGAGGTGCGAAAGCTTTGTTTCTTAAACCAT 679
DB 599 AATGTAGGTATGAACATTTTGGAGGTGTGAAGAGCTTTGTTCTCTTAAACCAT 658
QY 680 CTT-----AGTCTCTGCCACACTTGACACTCGCTCAAGTGAAGCAAGCAATAAGACC 733
DB 659 CTTAAGCACAATTTCTACGACCACTTGACATCTGTCAAGCAAGCAAGCAACTGCAGCC 718
QY 734 AACTCGGTGGAATAATTAATGTTTATGTAATAAATAA 773
DB 719 AGCTGCCATGAAAAATGTT--TGTTTATGGAATAAATAA 756
```

RESULT 4
AW555045/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AW555045
L0249C06-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA
clone L0249C06 3', mRNA sequence.
AW555045
AW555045.1 GI:7200468
EST.
house mouse.
Mus musculus

REFERENCE
AUTHORS

1 (bases 1 to 672)
Tanaka,T.S., Jaradat,S.A., Ljm,M.K., Kargul,G.J., Wang,X., Grabovac
M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)

JOURNAL
MEDLINE
COMMENT

20381348
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdnaelgusn.grc.nia.nih.gov
Plate: L0249 row: C column: 06
Seq primer: -21M13 Forward
High quality sequence stop: 672
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..672
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaEST:L0249C06-3"
/db_xref="taxon:10090"
/clone="L0249C06"
/clone_lib="NIA Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/note="vector: pSPORT1 (Gibco/BRL Life Technology);
Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
7 Newborn Ovary. The double-stranded cDNA was synthesized
by Gibco's kit with an oligo(dT) primer [NotI
primer-adaptor from Gibco/BRL]
[5'-pGACTAGTTCTAGATCGGCGGCCCTTTTTTTTTTTT-3'] from
2.5ug of total RNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker LL-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Yulan Piao."

BASE COUNT
ORIGIN

195 a 153 c 105 g 219 t
Query Match 61.1%; Score 476.4; DB 10; Length 672;
Best Local Similarity 86.4%; Pred. No. 2.3e-109;
Matches 577; Conservative 0; Mismatches 76; Indels 15; Gaps 4;
QY 112 TCAGACCCACGCGGAGCAGCTCTTCAGTGAAGAGGAAGCAATCGGAGGTCAAGCAAT 171
DB 672 TCAGACCCACGCGAGC-AGCAGCTCTTCGCTGAAGAGGAAGCAATCGGAGGTCAAGCAAT 614
QY 172 GAACGTGGAGCATGAGTTAACTCTCTCGTGGTGGAGGAATTCATCTCTGGTGTCCAAAAA 231
DB 613 GAATGTGGAACATGAAGTTAACTCTCTCGTGGTGGAGGAATTCATCTCTGGTGTCCAGAAA 554

Db	402	TTGATTTGACATTTGTTATTCGTCGAAGATAAATGTTGGTTTGCAGATCTCGGGGTATCTGGT	343
QY	445	AAACTGGAATAATTAAGTTAAAGGACAAACATGAAGTTCCCTTATGTTATTTTATAGACCT	504
Db	342	AAACTGGAATAATTAAGTTAAAGGACAAACATGAAGTTCCCTTATGTTATTTTATAGACCT	283
QY	505	TTGTTAAACAAAGGGGACTTCTTGAGAGTCCTGTTTTTATACCTTGAGGACAAACATTA	564
Db	282	TTGTTAAACAAAGGGGACTTCTTGAGAGTCCTGTTTTTATACCTTGAGGACAAACATTA	223
QY	565	CAATGTTAAATAAACAAAACCTGTTATTTTCTTAAAGAGGTAATCGGGAGACG	624
Db	222	CAATGTTAAATAAACAAAACCTGTTATTTTCTTAAAGAGGTAATCGGGAGACG	163
QY	625	TAGGCAATAAAATGTTTTCAGAGGTGCGAAAAGCTTTTGTCTTAAACCATCTTTAG	684
Db	162	TAGGCAATAAAATGTTTTCAGAGGTGCGAAAAGCTTTTGTCTTAAACCATCTTTAG	103
QY	685	TCCTGCGACACTTGACACTCCGTCAAAGTGAGAGCGAACTAAAGACCAACTCGGTGG	744
Db	102	TCCTGCGACACTTGACACTCCGTCAAAGTGAGAGCGAACTAAAGACCAACTCGGTGG	43
QY	745	AAATATATTATGTTTATGTAATAAAAAAAATCATGT	780
Db	42	AAATATATTATGTTTATGTAATAAAAAAAATCATGT	7
RESULT	3		
LOCUS	AK013984		
DEFINITION	Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched library, clone:3110003A17:homolog to PRO2013, full insert sequence.		
ACCESSION	AK013984		
VERSION	AK013984.1	GI:12851586	
KEYWORDS	HTC: CAP trapper.		
SOURCE	Mus musculus (Strain:C57BL/6J) 13 days embryo head cDNA to mRNA, clone_11b:RIKEN full-length enriched mouse cDNA library		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Hatanai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakauchi, S., Ikegami, T., Kashiwagi, K., Fujii, S., Inoue, K., Ogawa, Y., Izawa, M., Ohara, E., Watanabe, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipillar sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	4		
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,		

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2003, 21:37:23 ; Search time 2228 seconds

(without alignments)
5669.874 Million cell updates/sec

Title: US-09-648-310-1

Perfect score: 780

Sequence: 1 ggcacgagctctctctgctcc.....gtaataaaaaaaatcatgt 780

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estro:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	646	82.8	690	14	BQ192542
c 2	636	81.5	642	9	AA891725
c 3	539.2	69.1	769	11	AK013984
c 4	476.4	61.1	672	10	AW555045
5	469.4	60.2	600	12	BG804701
6	456.8	58.6	551	13	BI791389

7	455	58.3	593	9	AA544950	vk38e06.r
8	452	57.9	669	10	BE687862	BE687862 uw52c09.x
9	438.8	56.3	613	9	AA607954	AA607954 vm41a03.r
10	425	54.5	591	9	AA542157	AA542157 vj80b03.r
11	420.4	53.9	507	9	AA008262	AA008262 mg77901.r
12	415.6	53.3	547	14	BQ445291	BQ445291 UI-M-ERO-
13	414.6	53.2	504	10	BE689324	BE689324 uw52c09.y
14	413.6	53.0	501	12	BF468668	BF468668 UI-M-BH3-
15	406.6	52.1	580	13	BM122169	BM122169 L0506D12-
16	402.6	51.6	481	9	AI154666	AI154666 ud37d02.r
17	400.6	51.4	511	10	AW823811	AW823811 uf59h02.y
18	392	50.3	471	9	AI786148	AI786148 uf1e09.y
19	390	50.0	465	9	AA220360	AA220360 mv70d02.r
20	387.2	49.6	564	13	BI791071	BI791071 id05d03.x
21	381.4	48.9	457	10	BB849249	BB849249 BB849249
22	373.6	47.9	430	9	AA816139	AA816139 vP44c07.r
23	372.4	47.7	501	9	AA986660	AA986660 uc82e12.y
24	370	47.4	490	9	AA833425	AA833425 ub58d03.r
25	368.4	47.2	437	9	AA462630	AA462630 vf90f10.r
26	357.6	45.8	564	9	AA712021	AA712021 vu60d08.r
27	347.8	44.6	521	9	AA986390	AA986390 uc82e12.x
28	337.8	43.3	347	9	AI234233	AI234233 EST230921
29	335.2	43.0	845	11	AF161398	AF161398 Homo sapi
30	331	42.4	384	13	BM390524	BM390524 UI-R-CN1-
31	320.4	41.1	876	11	AF116682	AF116682 Homo sapi
32	314.8	40.4	797	13	BI859694	BI859694 603386283
33	313.2	40.2	425	9	AA986641	AA986641 uc82d03.y
34	308.8	39.6	365	10	BB844309	BB844309 BB844309
35	301	38.6	749	14	BQ015110	BQ015110 UI-H-ED1-
36	299.4	38.4	755	12	BG530613	BG530613 602560008
37	294.8	37.8	573	13	BI110495	BI110495 602896473
38	294.8	37.8	791	12	BG528919	BG528919 602579350
39	293.8	37.7	752	10	AV716363	AV716363 AV716363
40	293.4	37.6	747	12	BG502291	BG502291 602550569
41	291	37.3	304	14	H34607	H34607 EST111677 R
42	288	36.9	634	9	AI133406	AI133406 HA2013 Hu
43	287.6	36.9	462	10	AW492391	AW492391 UI-M-BH3-
44	287.6	36.9	884	12	BG164526	BG164526 602342067
45	286.8	36.8	678	10	AW027019	AW027019 wt71b02.x

ALIGNMENTS

RESULT 1						
LOCUS	BQ192542/c	690 bp	mrna	linear	EST 30-APR-2002	
DEFINITION	UI-R-DRI-ckz-m-14-0-UI.sl UI-R-DRI Rattus norvegicus cDNA clone					
ACCESSION	BQ192542					
VERSION	BQ192542.1	GI:20368093				
KEYWORDS	EST.					
SOURCE	Norway rat.					
ORGANISM	Rattus norvegicus					
REFERENCE	1 (bases 1 to 690)					
AUTHORS	Bonaldi,M.F., Lennon,G. and Soares,M.B.					
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery					
JOURNAL	Genome Res. 6 (9), 791-806 (1996)					
MEDLINE	97044477					
COMMENT	Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoaresblue.weeg.uiowa.edu The sequence contained an oligo-dr track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A					

PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.1%; Score 109.6; DB 21; Length 657;
Best Local Similarity 67.5%; Pred. No. 1.1e-20;
Matches 154; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY	169	AATGAACGTGGAGCATGAGGTTAACCTCCTGGTGGAGGAAATTCATCGTCTGGGTTCCAA	228
DB	198	AATCAACGTAGACGACGAGAGATTTCAGAACTGGAAGAAGAGATCCATCGTCTGGTCTCG	257
QY	229	AAATGCCGATGGGAACTGAGTGTGAAGTTGGGGTCTCTTCCAGACGACAGATGTGC	288
DB	258	TCAGACGATGGGCTCTTACAAAGGTGACGTTGGAGTGTGTTCATGACGATCGATGTGC	317
QY	289	CAATCTCTTTGAAGCGTTGGTGGGAACTCTGAAAGCCGCAACGAAAGGAGATTTGTAC	348
DB	318	CAACATCTTTGAAGCATTTGGTGGGACTCTGAGAGCTGCCAAGAAACGCAAAATAGTCGC	377
QY	349	GTACGACGAGAGCTGCTTTTGAAGGTGTTCATGATGATGTTGACAT	396
DB	378	ATTTGAAGGTGAACCTTTTGCTTCAAGGCGTTTCAGATAAGGTTGAGAT	425

Search completed: January 16, 2003, 21:51:13
Job time : 307 secs

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 02-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158569.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

Db	146	GCCATGACGTGGAGGAGGAGGTCGGGAGGCTCAAGGAGGAGATCCATAGGCTCGGCCAG	203
Qy	227	AAAAATGCCGATGGGA ^{AA} ACTGAGTGTGAAGTTTGGGCTCTCTTCCAAAGACGACAGATGT	286
Db	206	CAGCAGCCGATGGCTCTTACAAGGTCAAGTTTGGCGTTCTCTTCAACGACGATCGGTGT	265
Qy	287	GCCATCTCTTTGAAGCGTTGGTGGGA ^{CT} CTGAAAGCCGCA ^{AA} ACGAAGGAAGATTTGTT	346
Db	266	GCAATATCTTTTGAAGCACTAGTTGGCACCTTGAGGCGCCCAAGAGAGATCTTTG	325
Qy	347	ACGTACGCAGCAGAGCTGCTTTTTCGAAGGTCTTCATCATGATGTTCACATTTGATTTGCTG	406
Db	326	ACCTACGAAGCGAGCTGCTTCTCAAGGTGTCCATGACAACGTGGAGATAACCCCTGTTG	385
Qy	407	C	407
Db	386	C	386
RESULT 15			
Id	AAC53673 standard; DNA; 657 bp.		
Ac	AAC53673;		
XX	18-OCT-2000 (first entry)		
DT	Arabidopsis thaliana DNA fragment SEQ ID NO: 75315.		
DE	Hybridisation assay; genetic mapping; gene expression control;		
KW	protein identification; signal transduction pathway;		
KW	metabolic pathway; promoter; termination sequence; ss.		
XX	Arabidopsis thaliana.		
XX	EP1033405-A2.		
XX	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
XX	25-FEB-1999; 99US-0121825.		
PR	05-MAR-1999; 99US-0123180.		
PR	09-MAR-1999; 99US-0123348.		
PR	23-MAR-1999; 99US-0125788.		
PR	25-MAR-1999; 99US-0126264.		
PR	29-MAR-1999; 99US-0126785.		
PR	01-APR-1999; 99US-0127462.		
PR	06-APR-1999; 99US-0128234.		
PR	08-APR-1999; 99US-0128714.		
PR	16-APR-1999; 99US-0129845.		
PR	19-APR-1999; 99US-0130077.		
PR	21-APR-1999; 99US-0130449.		
PR	23-APR-1999; 99US-0130510.		
PR	23-APR-1999; 99US-0130891.		
PR	28-APR-1999; 99US-0131449.		
PR	30-APR-1999; 99US-0132048.		
PR	04-MAY-1999; 99US-0132407.		
PR	05-MAY-1999; 99US-0132484.		
PR	06-MAY-1999; 99US-0132485.		
PR	06-MAY-1999; 99US-0132486.		
PR	07-MAY-1999; 99US-0132487.		
PR	11-MAY-1999; 99US-0134256.		
PR	14-MAY-1999; 99US-0134218.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134221.		
PR	18-MAY-1999; 99US-0134370.		
PR	19-MAY-1999; 99US-0134768.		
PR	20-MAY-1999; 99US-0134941.		
PR	21-MAY-1999; 99US-0135124.		
PR	24-MAY-1999; 99US-0135353.		
PR	24-MAY-1999; 99US-0135629.		

Qy	87	AGGCGCGGAAGCTGAGTGCAGGGTTCAGACCCACGCGGCGCAGCGTCTTTCAGTGAAGA	146
Db	90	AGTGCACCTGAAGAAGGTAAACGGGTCTCCAGACCCACGCGCGCCAGTTCCTCCGCGGGAAG	149
Qy	147	AGGAAGCAACTCGGAGGCTACGAATGAACCTGGAGCATGAGGTTAACTCTCTGGTGGAGG	206
Db	150	GAAAAACCGCGCAGAGAGGCGCAGCAATGAATCTGGATCACGAGGTTAACCTCTTAGTGGAGG	209
Qy	207	AAATTCATCGTCTGGGTTCCAAAATATGCCGATGGGAAACTGAGTGTGAAGTTTGGGGTCC	266
Db	210	AAATTCATCGTTTGGGTTCAAAAATATGCTCATGCGAAAGTTAAACGGTGAAGAAATTTGGGGTTC	269


```

Db 193 TCAGGAATCTGGTTAGGAATTCAGCAATGAGATTTTTCGGGGCAAGGATGGGAATG 134
QY 663 TGTTTCTTAAACCAATCTTAGTCT--CTGCCACACTTGACACTCCGTCAAAGTGAGAAAG 720
Db 133 TTGTTTCATAAATAATAGACATTTCTATAGATATATTGACATTTCTCGAAAGCAACAAG 74
QY 721 CGAACTAAAGACCACTCGGGTCGAAATAATTATGTTATGTAATAAAAAA 773
Db 73 CAACATGAAGACCACTCTATGAGAAATATTATGATGTTTATGTAATAAGA 21

RESULT 10
AAZ21551
ID AAZ21551 standard; DNA; 177 BP.
XX
AC AAZ21551;
XX
DT 01-DEC-1999 (first entry)
XX
DE Fragment of progression suppressed gene 13 (PSGen13).
XX
KW Progression suppressed gene; PSGen; progression elevated gene; PEGen;
KW tumour; reciprocal subtraction differential RNA display; RSDD;
KW differential expression; gene cloning; cancer; ss.
XX
OS Rattus sp.
XX
PN WO9943844-Al.
XX
PD 02-SEP-1999.
XX
PE 26-FEB-1999; 99WO-US04323.
XX
PR 27-FEB-1998; 98US-0032684.
XX
PR 03-NOV-1998; 98US-0185115.
XX
PR 23-NOV-1998; 98US-0197889.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Fisher PB;
XX
WPI: 1999-550872/46.
XX
Identifying nucleic acids differentially expressed between two samples,
particularly sequences involved in tumour progression
XX
Examples; Fig 23; 110pp; English.
XX
This sequence is a fragment of the progression suppressed gene 13
(CC (PSGen13)). This gene has suppressed expression in progressed tumour
cells. PSGen13 was identified using new methods for identifying nucleic
acids differentially expressed between two samples. The method involves
performing reciprocal subtraction differential RNA display (RSDD) between
the two samples to generate two subtraction samples. The subtraction
samples are amplified and compared to identify those nucleic acids that
are differentially expressed. The method is used to identify and clone
differentially expressed genes, particularly those with increased or
reduced expression during tumour cell progression, e.g. progression
of suppressed genes (PSGen) and progression elevated genes (PEGen). The
method reduces the complexity of the band pattern produced in
conventional differential RNA display (where bands may be obscured,
resulting in false positive signals) since most bands common to both
samples are eliminated, allowing identification and cloning of genes
displaying anticipated differential expression. RSDD requires only a
single anchored primer for amplification and reamplified cDNA can be
analysed by reverse Northern blotting.
XX
Sequence 177 BP; 72 A; 28 C; 30 G; 47 T; 0 other;
Query Match 20.1%; Score 157; DB 20; Length 177;
Best Local Similarity 100.0%; Pred. No. 2.5e-34;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

```

QY 624 GTAGGCAATAAATCTTTTCAGAGGTGCGAAAAAGCTTTTGTCTTTAAACCATCTTA 683
Db 1 GTAGGCAATAAATCTTTTCAGAGGTGCGAAAAAGCTTTTGTCTTTAAACCATCTTA 60
QY 684 GTCCTCTGCCACACTTGACACTCCGTCAAAGTGAGAGCGAACTAAAGACCAACTCGG 743
Db 61 GTCCTCTGCCACACTTGACACTCCGTCAAAGTGAGAGCGAACTAAAGACCAACTCGG 120
QY 744 GAAATATTTATGTTTATGTAATAAAAAAATCATGT 780
Db 121 GAAATATTTATGTTTATGTAATAAAAAAATCATGT 157

RESULT 11
AAZ25869
ID AAZ25869 standard; cDNA; 177 BP.
XX
AC AAZ25869;
XX
DT 08-JUN-1999 (first entry)
XX
DE Clone PSGen-13 sequence.
XX
KW Isolation; identification; reciprocal subtraction differential display;
KW RSDD; gene expression; ss.
XX
OS Mus sp.
XX
PN US5882874-A.
XX
PD 16-MAR-1999.
XX
PE 27-FEB-1998; 98US-0032684.
XX
PR 27-FEB-1998; 98US-0032684.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Fisher PB;
XX
WPI: 1999-214060/18.
XX
Identifying differentially expressed nucleic acids between two
samples - by reciprocal subtraction
XX
Examples; Fig 23; 29pp; English.
XX
The invention relates to a method of isolating and identifying
differentially expressed nucleic acids between two samples by using
a reciprocal subtraction differential display (RSDD) method. This
sequence represents clone PSGen-13 which is a novel gene.
XX
Sequence 177 BP; 72 A; 28 C; 30 G; 47 T; 0 other;
Query Match 20.1%; Score 157; DB 20; Length 177;
Best Local Similarity 100.0%; Pred. No. 2.5e-34;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 GTAGGCAATAAATCTTTTCAGAGGTGCGAAAAAGCTTTTGTCTTTAAACCATCTTA 683
Db 1 GTAGGCAATAAATCTTTTCAGAGGTGCGAAAAAGCTTTTGTCTTTAAACCATCTTA 60
QY 684 GTCCTCTGCCACACTTGACACTCCGTCAAAGTGAGAGCGAACTAAAGACCAACTCGG 743
Db 61 GTCCTCTGCCACACTTGACACTCCGTCAAAGTGAGAGCGAACTAAAGACCAACTCGG 120
QY 744 GAAATATTTATGTTTATGTAATAAAAAAATCATGT 780
Db 121 GAAATATTTATGTTTATGTAATAAAAAAATCATGT 157

RESULT 12

```

PD 11-APR-2002.
 XX 02-OCT-2001; 2001WO-US30732.
 XX 02-OCT-2000; 2000US-237271P.
 XX (FARB) BAYER CORP.
 XX Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
 XX Thiagalingam A, Lewis ME;
 XX WPI; 2002-426115/45.
 DR New isolated nucleic acid that is differentially expressed in cancer
 XX tissues useful for determining the presence of colon cancer in a cell
 XX or tissue type, and in antisense therapy
 XX Claim 1; Fig 1; 796pp; English.
 XX
 CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
 CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
 CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
 CC used in antisense therapy. An antibody immunoreactive with a polypeptide
 CC encoded by (I) is useful for detecting cancer in a patient sample, and
 CC for detecting the presence or absence of a polynucleotide encoded by a
 CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived
 CC from (I) can be used for determining the presence of a nucleic acid which
 CC hybridises to (I), and for determining the phenotype of cells in a sample
 CC of cells from a patient. (I) is useful for determining the presence of
 CC colon cancer in a cell or tissue type, for determining the presence or
 CC state of other type of cancer, in antisense therapy, to generate
 CC macroarrays on a solid surface, to identify a chromosome on which the
 CC corresponding gene resides, and in tissue profiling, forensics, genetic
 CC analysis, mapping and diagnostic applications. (I) can be used to raise
 CC antibodies, and to screen for peptide analogues and antagonists.
 XX
 SQ Sequence 627 BP; 171 A; 151 C; 153 G; 147 T; 5 other;
 Query Match 27.6%; Score 215.6; DB 24; Length 627;
 Best Local Similarity 83.3%; Pred. No. 1.3e-50;
 Matches 245; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
 QY 147 AGGAAGCAATCGGAGGTCACGATGAACGTGGAGCAGTGAACCTCTGGTGGAGG 206
 Db 294 AAGAGGACCTACGATGCGAGCAATGAATGGATCAGAGCTTAACCTTATGTGGAGG 235
 QY 207 AAATTCATCGTCTGGTTCCTCAAAATGCGATGGAACTGAGTGTGAAGTTGGGTC 266
 Db 234 AAATTCATCGTCTGGTTCCTCAAAATGCGATGGAACTGAGTGTGAAGTTGGGTC 175
 QY 267 TCTTCCAGACGACAGATGTCCTTGAAGCGTTGGTGGAACTCTGAAAGCGC 326
 Db 174 TCTTCCGATGATGAATGTGCAACCTCTTTGAAGCATTGGTGAAGTCTTAAAGCTG 115
 QY 327 CAACACGAAGCAAGATTTGATGACGACGAGAGCTGCTTTGCAAGGTTTCATGATG 386
 Db 114 CAACACGAAGCAAGATTTGATGACATATCAGGAGCTCTCTTGAAGGTTTCATGATG 55
 QY 387 ATGTTGACATGTTATGTCGCAAGATTAAATGTTGTTTCAGATCTGGGGTATC 440
 Db 54 ATGTTGACATTAATATGTTGCAAGATTAAATGTTGTTTACATATCTTATGTACC 1
 RESULT 9
 ABK45388/C
 ID ABK45388 standard; cDNA; 552 BP.
 XX
 AC ABK45388;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE cDNA encoding colon tumour protein, SEQ ID No 939.
 XX

KW Human; colon tumour; vaccine; colon cancer; immunogenic;
 KW immunotherapy; gene; ss.
 OS Homo sapiens.
 XX WO200212328-A2.
 XX 14-FEB-2002.
 XX 31-JUL-2001; 2001WO-US24218.
 XX 03-AUG-2000; 2000US-223283P.
 PR 28-MAR-2001; 2001US-279763P.
 PR 29-JUN-2001; 2001US-302051P.
 XX (CORI-) CORIXA CORP.
 XX
 PI King GE, Meagher MJ, Xu J, Secrlist H;
 XX WPI; 2002-241739/29.
 XX New colon cancer polypeptides and polynucleotides, useful as vaccines,
 PT for diagnosing, preventing, and treating colon cancer, and as markers
 PT for the progression of cancer -
 XX
 PS Claim 1; SEQ ID No 939; 147pp; English.
 XX
 CC The invention relates to polynucleotides encoding colon tumour proteins.
 CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
 CC compositions, such as vaccines, for the diagnosis, prevention, and
 CC treatment of colon cancer. Polynucleotide sequences may be used as
 CC hybridisation probes or primers, and in the design and preparation of
 CC ribozyme molecules for inhibiting expression of tumour polypeptides and
 CC proteins in tumour cells. The compositions are useful for stimulating an
 CC immune response against cancer, particularly for the immunotherapy of
 CC colon cancer, and as markers for the progression of cancer.
 CC ABK44450-ABK46237 represent coding sequences of human colon tumour
 CC proteins of the invention.
 CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
 CC for this patent did not form part of the printed specification but was
 CC supplied by the European Patent Office.
 XX
 SQ Sequence 552 BP; 190 A; 104 C; 75 G; 183 T; 0 other;
 Query Match 26.5%; Score 206.6; DB 24; Length 552;
 Best Local Similarity 68.7%; Pred. No. 4.4e-48;
 Matches 366; Conservative 0; Mismatches 144; Indels 23; Gaps 5;
 QY 263 GTCTCTTCCAGACGACAGATGTCCTTTGAAGCGTTGGTGGAACTCTGAAA 322
 Db 552 GTCTCTTCCGTCATGATAAATGTGCAACCTCTTTGAAGCATTGGTAGAACTCTTAAA 493
 QY 323 GCGGCAAAACGAGGAGGATGTTAGCTAGCAGGAGAGCTGCTTTTGCAGGTGTCAT 382
 Db 492 GCTGCAAAACGAGGAGGATGTTAGCATATCCAGAGAGCTGCTTCTCAGAGTGTTCAT 433
 QY 383 GATGATGTTGACATTTGATGCTGCAAGATTAAATGTTGAGTGTGAGGTGTA---- 438
 Db 432 GATGATGTTGACATTTATATTTACTGCAAGATTAAATGTTGAGTGTGATCTTTTGTACTGC 373
 QY 439 -----TCTGTAACTGGAATAATAAGTTAAAGGACAAACAT---GAAGTTC 484
 Db 372 CATTTTTTGTCTGCTAAACTGGAAT-ATAAGTGAAGAACAAACATTTGAACATACT 314
 QY 485 TTATGTTATTTTATACCTTTGTAACAAAGGGGA--CTTGTGTGAGAGTCTCTGTTTT 542
 Db 313 TAATGTTATTTTATAGAACTTTGTAACGAAAGGAGATTCATGTTTGTAGAGCTGCTCT 254
 QY 543 TATACCTTGGAGCAAAACATTAACATGTAATAAATAACAAACCTGTTATTTTTC 602
 Db 253 TTTTATATCTTGAAAGAAATCTATGATGCTATAAATAAATAATCTTATTATTTC 194
 QY 603 TTAAGAAGGTAATCGGAGAGCTAGGCAATAAATGTTTTTTCAGAGGTGCGAAAAAGCTTT 662

DE	Human colon cancer related nucleotide sequence, SEQ ID NO:750.
XX	
KW	Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW	genetic analysis; diagnostic; antisense therapy; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200229086-A2.
PD	
PP	11-APR-2002.
XX	
XX	02-OCT-2001; 2001WO-US30732.
XX	
PR	02-OCT-2000; 2000US-237271P.
XX	
PA	(FARB) BAYER CORP.
XX	
PI	Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI	Thiagalingam A, Lewis ME;
XX	
DR	WPI; 2002-426115/45.
XX	
PT	New isolated nucleic acid that is differentially expressed in cancer
PT	tissues useful for determining the presence of colon cancer in a cell
PT	or tissue type, and in antisense therapy
PS	Claim 1; Fig 1; 796pp; English.
XX	
CC	ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC	expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
CC	encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC	used in antisense therapy. An antibody immunoreactive with a polypeptide
CC	encoded by (I) is useful for detecting cancer in a patient sample, and
CC	for detecting the presence or absence of a polynucleotide encoded by a
CC	nucleic acid which hybridises to (I) in a cell. A probe/primer derived
CC	from (I) can be used for determining the presence of a nucleic acid which
CC	hybridises to (I), and for determining the phenotype of cells in a sample
CC	of cells from a patient. (I) is useful for determining the presence of
CC	colon cancer in a cell or tissue type, for determining the presence or
CC	state of other type of cancer, in antisense therapy, to generate
CC	macroarrays on a solid surface, to identify a chromosome on which the
CC	corresponding gene resides, and in tissue profiling, forensics, genetic
CC	analysis, mapping and diagnostic applications. (I) can be used to raise
CC	antibodies, and to screen for peptide analogues and antagonists.
XX	
SQ	Sequence 594 BP; 157 A; 142 C; 134 G; 149 T; 12 other;
Query Match	33.7%; Score 263; DB 24; Length 594;
Best Local Similarity	76.0%; Pred. No. 5.6e-64;
Matches 323; Conservative	0; Mismatches 102; Indels 0; Gaps 0;
QY	16 CGTCCCTCCTTCACATGCGAGCCCTTTCTTAGCCGCAGACCATTCTTCTTCTGT 75
Db	
QY	45 CTGCTCTCTCCTCCACAGCGCTTCTTCTTGCCCTCAGCCACTTCTCTCTGGCC 104
Db	
QY	76 TGTTCTCCTCCTAGCGCGGCAAGCTGAGTGCAGGTTTCAGACCCAGCGGCCGAGCCTC 135
Db	
QY	105 TCACCCTCCCAGTGCAGTGAAGAAGTAACCGGTCAGACCCAGCGGCCAGCTTCT 164
Db	
QY	136 TTCAGTGAAGAAGAACAAATCGGAGGTCAGCAATGAACGCGGAGCATGAGTTAACT 195
Db	
QY	165 CCGCCGGGAAGAAAACCCGCGAGAGCGAGCAATGAATGTGGATCACGAGTTAACT 224
Db	
QY	196 CCTGGTGGAGAAATTCATCGTCTGGTTCCAAATAATCCGATGGAAACTGAGTGTGAA 255
Db	
QY	225 CTTAGTGGAGAAATTCATCGTCTGGTTCCAAATAATGCTGATGGAAGTGAAGCGTGA 284
Db	
QY	256 GTTTGGSGTCTCTCCAAGACACAGATGTGCCAATCTCTTTGAAGCGTCTGGTGGAA 315
Db	
QY	285 ATTTGGSGTCTCTCCGTGATGAATAATGTCACCTCTTTGAAGCATTTGGTAGGAC 344
Db	
QY	316 TCTGAACCGCCAAACGAAGGAAGATTTGCTAGCTACGAGGAGACTGCTTTTTGCAAG 375
Db	

Db 345 TCTTAAAGCTCAAAACGAAGGAGATTGTAACAATCCAGGAGAGCTGCTTCTTCAAGG 404

QY 376 TGTTCATGATGATGTTGACATTGCTGCAAGATTAAATGTGGTTGACAGATCTGGG 435
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 405 NGTTCATGATGATGTTGACATTATATTACTGCAAGATAATNGGGTTACATATCTTAT 464

QY 436 GTATC 440
| | | | |

Db 465 GTACC 469

RESULT 6

AAC66413

ID AAC66413 standard; DNA; 712 BP.

XX AAC66413;

XX 14-FEB-2001 (first entry)

XX Human secreted protein coding sequence SEQ ID NO: 14.

DE

XX Cytostatic; Immunosuppressive; nootropic; neuroprotective; antiviral;
anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder
neurological disease; infection; human; secreted protein; ss.

OS Homo sapiens.

PN WO2000058350-A1.

XX 05-OCT-2000.

PF 22-MAR-2000; 2000WO-US07483.

PR 26-MAR-1999; 99US-0126596.
22-DEC-1999; 99US-0171552.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

PI WPI; 2000-602357/57.

DR P-PSDB; AAB32005.

XX Nucleic acid molecules encoding human secreted proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -

XX Claim 1; Page 336; 423pp; English.

XX The invention relates to the isolation of genes AAC66410-C66458 encoding 49 human secreted proteins AAB32002-B32050. The genes can be used to generate fusion proteins by linking to the gene for the human immunoglobulin G Fc portion (SEQID1) for increasing the stability of the fusion protein as compared to the human protein only. The genes and proteins are useful for preventing, ameliorating or treating medical conditions, e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

XX Sequence 712 BP; 241 A; 121 C; 149 G; 201T; 0 other;

Db 343 TCCTTAAGCTGCAAAAGGAAGATGTAACATATCCAGGAGAGCTGCTTCTGCAAGG 402
 QY 376 TGTTTCATGATGTTGACATTTGATTCGCAAGATTAATGTTGAGATCTGGGG 435
 Db 403 TGTTTCATGATGTTGACATTTGATTCGCAAGATTAATGTTGAGATCTGTTAT 462
 QY 436 GTA-----TCTGTAACCTGGAATTAATGTTGAGATCTGTTAT 477
 Db 463 GTACTGCCATTTTGTGTTCTGTAACCTGGAAT-ATAAAGTGAAGAACAACATTTGA 521
 QY 478 AAGTTCCTTATGATTTTATAGACCTTTGTAACAAAGGGGACT--TCTTGAGAGTC 535
 Db 522 ACATCTTATGATTTTATAGACCTTTGTAACAAAGGGGAGATTCATGTTTAGAAGT 581
 QY 536 CTGTTTTTATACCTTGAGCAAAATTTACAATGTAAAAATAAACAACTGTTATTTT 595
 Db 582 CTGCTCTTTTATATCTTGAAAGAAATCTATGATGATGCTATATAAATAATCTTAT 641
 QY 596 TTTTTCCTTAAGAGTAATCGGAGACGATGAGCAATAAATGTTTACAGAGTGGGAAA 655
 Db 642 ATTTTCTCAGGAATCTGTTTGAATATGCAAGCAATGAGATTTTTCGGGGCAGGGAT 701
 QY 656 AAGCTTTTGTGTTTCTTAACCATCTTAGTCT--CTGCCACACTTGACACTCCGTCAGAG 713
 Db 702 GGAATGTTGTTCAATAATAATAGACATTTTCTATAGATATTTGACATCTCGCAAG 761
 QY 714 TGAGAACGCACTTAAGACCAACTCGGGTGGAAATATTTATGTAATAAAAAAAA 773
 Db 762 CAACAAGCAACTGAAGACCAACTCTATGAGAAATATTTATGTTTATGTAATAAGA 821

RESULT 4

AAFI6110
 ID AAFI6110 standard; cDNA; 778 BP.
 AC AAFI6110;
 XX
 DT 13-MAR-2001 (first entry)
 DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:545.
 XX
 KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO20005174-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05988.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-587513/55.
 DR P-PSDB; AAB56907.
 XX
 PT Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -
 XX
 PS Claim 1; Page 1004-1005; 2338pp; English.
 XX
 CC AAFI5566 to AAFI6505 encode the human prostate cancer associated

CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
 CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders, wounds, and infectious diseases. AAFI6506 to AAFI6514 to
 CC AAB57303 represent sequences used in the exemplification of the present
 CC invention.
 XX

Sequence 778 BP; 250 A; 151 C; 148 G; 221 T; 8 other;

Query Match 36.9%; Score 288.2; DB 21; Length 778;
 Best Local Similarity 69.9%; Pred. No. 5e-71;
 Matches 451; Conservative 2; Mismatches 167; Indels 25; Gaps 4;

QY 16 GGTCCCTCCCTTCCCACTGCAGCTTCTTAGCCCGAACCACCTTCTTCTCTGCT 75
 Db 32 CCTGCTCTCTCTCCACAGCGCTTCTTCTTGGCTCAGCCACTTCTTCTCTCGCC 91
 QY 76 TGTTCCTCCTTAGCGCGGAGCTGAGTGCAGGGTTTCAGACCCACGCGGCGCAGCCTC 135
 Db 92 TCACCTCCCTCAGTGCATGAGAGGTAACCGGGTCCAGACCCACGCGGCGCAGTCT 151
 QY 136 TTAGTGAAGAGGAGCAATCGGAGGTCAGCAATGAACGTGGAGCATGAGTTAACT 195
 Db 152 CCGCGGGGAGGAGAAACCGCGCAGAGGAGCAATGAATGTGGATCAGGTTAACT 211
 QY 196 CCTGTGAGGAAATTCATCTCTGGGTTCCAAAATCCGATGGGAACATGAGTGTGA 255
 Db 212 CTTAGTGAGGAAATTCATCTTGGGTTCAAAAAATCTGATGGAAGTTAAGCGTGA 271
 QY 256 GTTTGGGTCTCTTCCAAAGACAGATGTGCCAATCTTTTGAAGCGTTGGTGGGAAC 315
 Db 272 ATTTGGGTCTCTTCCGTGATGATAATGTGCCAATCTTTGAGCATTTGGTAGAAC 331
 QY 316 TCTGAAGCCCAAAACGAGAGATTTGATGACGAGAGAGAGTGTGTTTGAAGG 375
 Db 332 TCTTAAAGCTGCAAAACGAGAGATTTGAACATATCCAGGAGAGCTGCTTCTCAAGG 391
 QY 376 TGTTCATGATCATCTGACATTTGATGTCGCAAGATTAATGTTGAGATCTGGGG 435
 Db 392 TGTTCATGATCATCTGACATTTATTTACTGCAAGATTAATGTTGATATATCTTAT 451
 QY 436 GTA-----TCTGTAACCTGGAATTAATGTTAAGGACAAACAT---G 477
 Db 452 GTACTGCCATTTTGTGTTCTGTTAACTGGAAT-ATAAGTGAAGAACAAACATTTGA 510
 QY 478 AAGTTCCTTATGATTTTATAGACCTTTGTAACAAAGGGGGA-----CTTGTGAGA 531
 Db 511 ACATCTTATGATTTTATAGAACTTTGTAACGAGGAGAGATTCATGTTTAGAAGT 570
 QY 532 AGTCTGTTTTTATACCTTGAGCAAAACATTTACAATGTAATAAATAACAAACCTGTTA 591
 Db 571 CTGCTCTTTTATATCTTGAAGAAATCTATGATGATGCTATATAAATAATCTTAT 630
 QY 592 TTTTCTTTTCTTAAAGAGGTAATCGGAGAGAGTGGCAATAAAAA 636
 Db 631 ATTTCTMAGNATMTGGTTGANATCTCGGAAAGCAACAGCAAA 675

RESULT 5

ABQ57055
 ID ABQ57055 standard; cDNA; 594 BP.
 XX
 AC ABQ57055;
 XX
 DT 02-AUG-2002 (first entry)
 XX

Query Match 96.8%; Score 754.8; DB 20; Length 800;
Best Local Similarity 99.5%; Pred. No. 1.3e-202;
Matches 778; Conservative 0; Mismatches 2; Indels 2;

QY	1	GGCAGGAGCTCTCTCGTGCCTCCCTCTCCACTGCAGCCTTCTCTTTAGCCGAACCA	60
Db	1	GGCAGAGCTCTCTCGTGCCTCCCTCTCCACTGCAGCCTTCTCTTTAGCCGAACCA	60
QY	61	CTTCCTTCTTCTGCTGTGTTCTCCTTAGGCGCGGAAGCTGAGTGCAGGGTTTCAGACCCA	120
Db	61	CTTCCTTCTTCTGCTGTGTTCTCCTTAGGCGCGGAAGCTGAGTGCAGGGTTTCAGACCCA	120
QY	121	CGCGGCGAGCAGCTCTTCAGTGAAGAAGCAACATCGAGGGTCAGCAATGAACGTGGA	180
Db	121	CGCGGCGAGCAGCTCTTCAGTGAAGAAGCAACATCGAGGGTCAGCAATGAACGTGGA	180
QY	181	GCATGAGGTTAACTCTCGTGTGAGGAAATTCATCGTCTGGTTCACAAAATGCCGATGG	240
Db	181	GCATGAGGTTAACTCTCGTGTGAGGAAATTCATCGTCTGGTTCACAAAATGCCGATGG	240
QY	241	GAAACTGAGTGAAGTTTGGGGTCTCTCCAAAGACGACAGATGTGCCAATCTCTTTTG-	299
Db	241	GAAACTGAGTGAAGTTTGGGGTCTCTCCAAAGACGACAGATGTGCCAATCTCTTTTGA	300
QY	300	AAGCGTTGGTGGGAACCTCTGAAG-CCGCCAAAACGAAGAAAGATTGTTACGTACGCAGCA	358
Db	301	AACCGTTGGTGGGAACCTCTGAAGCCGCAACGAAGGAAGATTGTTACGTACGCAGCA	360
QY	359	GAGCTGCTTTTGCAGAGTGTTCATGATGATGTTGACATGTTATTTGCTGCCAAGATTAAATGT	418
Db	361	GAGCTGCTTTTGCAGAGTGTTCATGATGATGTTGACATGTTATTTGCTGCCAAGATTAAATGT	420
QY	419	GGTTTGCAGATCTGGGGTATCTGTAACCTGGAATTAATTAAGTTAAAGGACAAACATGA	478
Db	421	GGTTTGCAGATCTGGGGTATCTGTAACCTGGAATTAATTAAGTTAAAGGACAAACATGA	480
QY	479	AGTTCCCTATGTAATTTTATAGACCTTTGTAACAAAGGGGACTGTGTTGAGAAAGTCCTG	538
Db	481	AGTTCCCTATGTAATTTTATAGACCTTTGTAACAAAGGGGACTGTGTTGAGAAAGTCCTG	540
QY	539	TTTTTATACCTTGGAGCAAAACATTACAATCTAAAAATAACAAACCTGTTATTTT	598
Db	541	TTTTTATACCTTGGAGCAAAACATTACAATCTAAAAATAACAAACCTGTTATTTT	600
QY	599	TTTCTTAAGAAAGTAAATCGGGAGAGCTAGGCAATATAATGTTTTTCAGAGGTCCGAAAAAG	658
Db	601	TTTCTTAAGAAAGTAAATCGGGAGAGCTAGGCAATATAATGTTTTTCAGAGGTCCGAAAAAG	660
QY	659	CTTTTGTGTTTCTTAACCATCTTAAAGTCTCGCCACACTTGACACTCCGTCAAAAGTGAGA	718
Db	661	CTTTTGTGTTTCTTAACCATCTTAAAGTCTCGCCACACTTGACACTCCGTCAAAAGTGAGA	720
QY	719	AGCGAACTAAAGACCAACTGCGGTGGAAAAATTTATGTTTATGTAATAAAAAAAATCAT	778
Db	721	AGCGAACTAAAGACCAACTGCGGTGGAAAAATTTATGTTTATGTAATAAAAAAAATCAT	780
QY	779	GT 780	
Db	781	GT 782	

RESULT 3
ABK11086
ID ABK11086 standard; cDNA; 835 bp.
XX
AC ABK11086;

XX
DT 05-JUN-2002 (first entry)

XX
DE Human Progression Supervisor

DE Human Progression Suppressed Gene 13 (HuPSGen 13), cDNA.

XX Human; Progression Suppressed Gene 13; rPSGen 13; cytostatic; cancer;
KW

KW	blood vessel; nasopharyngeal tumour; thyroid tumour; leukaemia;
KW	lymphoma; breast; lung; prostate; ovary; colon; gene; ss.
XX	
OS	Homo sapiens.

PH	Key	Location/Qualifiers
CDS		197..442
FT	/*tag= a	
FT	/product= "Progression suppressed gene 13 protein	
XX		
XX	W0200216419-A2.	
XX		
PD	28-FEB-2002.	
XX		
PF	27-AUG-2001; 2001WO-US26795.	
XX		
PR	25-AUG-2000; 2000US-0648310.	
XX	(UYCO) UNIV COLUMBIA NEW YORK.	
PA		
XX	Fisher PB, Kang D, Su Z;	
PI		
XX	WPI; 2002-280914/32.	
DR	P-PSDB; AAU76533.	
DR		

PT New rat and human Progression Suppressed Gene 13 for preventing the
PT growth of cancer cells and/or new blood vessels, and for treating
PT patients suffering from a cancer -

PS Claim 4; Fig 2; 53pp; English.

The invention relates to novel isolated nucleic acids which encode a rat or human Progression Suppressed gene 13 (PSG13) protein. The nucleic acids are useful for preventing the growth of cancer cells and/or new blood vessels, and for treating patients suffering from a cancer, e.g. nasopharyngeal tumour, thyroid tumour, leukaemia, lymphoma, or cancer of the breast, lung, prostate, ovary or colon. PSG13 may also be used to suppress the transformed phenotype of a malignant cell. Administration of PSG13 gene or protein may result in a decrease in tumour mass, number of cancer cells, serum tumour marker, tumour metastasis, vascularisation, perfusion, or rate of tumour growth, improved clinical symptoms, and/or increased patient survival. The present sequence represents the coding sequence of human Progression Suppressed Gene 13 (HupSGen 13).

Sequence 835 BP; 246 A; 160 C; 176 G; 253 T; 0 other;

Query Match	Score	DB	Length
Best Local Similarity	43.0%	335.2	835
Matches 539; Conservative	69.1%	Pred. No. 2.9e-84;	
0; Mismatches 218; Indels 23; Gaps 5			

Qy	16	GTCCCTCCCTTCTC	ACATG	CAGCC	TTTCT	CTATTAG	CCCGA	ACAC	ACATCTCTCTCT	TCGTGCT	75
Db	43	CCTGCTCCTCTCT	CACAG	CCCTCT	TTCTTTG	CTCAGG	CACATCT	CTCTTGCC	CT	102	
Qy	76	TGTTTCTCCCT	TAGGG	CGGA	AGCTG	ATG	TGACG	GGTTT	CACAC	CCACGCGG	135
Db	103	TCACCTCTCC	CACTG	CACTG	AAAG	GTAA	CCGGG	TCCAG	CCACCGCGG	CGCAGTTCT	162
Qy	136	TTCACTGA	GAAG	AAGCA	CAATCGG	AGGGT	CAGCA	ATGAAC	CTGGAG	CATGAGGTTAA	195
Db	163	CCGGCGG	GAAG	AAAA	CCGGC	AGAGG	CGACAA	TGAATCTG	GATC	CACGAGTTAA	222
Qy	196	CCTGGTGG	GAATTC	ATCGT	CTGG	TTCC	AAAAATG	CCGATG	GGAA	CTGAGTGTG	255
Db	223	CTTAGTGG	AGAAAT	TCATCGT	TTGG	TTTCA	AAAAATG	CTGATG	AAAGTTAA	ACCGGTGA	282
Qy	256	GTTTGGG	GTCTCT	TCCA	AGAC	ACATG	TGCC	AAATCTCT	TGAA	GCCTTGTGG	315
Db	283	ATTTGGG	GTCTCT	CGT	GAT	ATAAT	GTGCC	AACTCT	TTTGA	GCATTTGG	342
Qy	316	TCTGAA	AGCG	CAAA	CGAA	GAGAT	TGTT	TACG	TACG	AGAGAGCTGCT	375

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL: U13194; AAA68468.1; -
 DR EMBL: X83413; CAA58369.1; -
 DR InterPro: IPR003840; Herpes_Helicase.
 DR Pfam: PF02689; Herpes_Helicase; 1.
 KW DNA replication; ATP-binding; Helicase.
 FT NP_BIND 90 97 ATP (BY SIMILARITY).
 SQ SEQUENCE 824 AA; 93286 MW; 5A5507544E06184C CRC64;

Query Match 14.8%; Score 60; DB 1; Length 824;
 Best Local Similarity 23.6%; Pred. No. 40;
 Matches 21; Conservative 14; Mismatches 30; Indels 24; Gaps 3;

QY 1 MNVEHEVNLVEEHRIGSKNADKGL-----SVKFGVLFQDDRCANL----- 42
 Db 19 MSSAPKTELIVDKVASLSERLEGLPDMFRHMDPETEFSEFADALCIGIDEFAQPL 78
 QY 43 ----FEALV--GTLKAARRKRVITYAGEL 65
 Db 79 FLPLFKALLVTGTAGACKTNSIQTLAANL 107

RESULT 15

ID CALM_YEAST STANDARD; PRT; 147 AA.
 AC P06787;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-JUN-1988 (Rel. 06, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Calmodulin (Cam).
 GN CM1 OR YBR109C OR YBR0904.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95208357; PubMed=7900426;
 RA Davis T.N., Urdea M.S., Maslarsz F.R., Thorner J.;
 RT "Isolation of the yeast calmodulin gene: calmodulin is an essential
 RT protein.";
 RL Cell 47:423-431(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=95208357; PubMed=7900426;
 RA Mannhaupt G., Stucka R., Ehmlé S., Vetter I., Feldmann H.;
 RT "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
 RL Yeast 10:1363-1381(1994).
 RN [3]
 RP SEQUENCE OF 15-30 AND 128-145.
 RX MEDLINE=87228267; PubMed=3295478;
 RA Davis T.N., Thorner J.;
 RT "Isolation of the yeast calmodulin gene using synthetic
 RT oligonucleotide probes.";
 RL Meth. Enzymol. 139:248-259(1987).
 RN [4]
 RP CALCIUM-BINDING DATA.
 RX MEDLINE=93278279; PubMed=1304352;
 RA Brockerhoff S.E., Edmonds C.G., Davis T.N.;
 RT "Structural analysis of wild-type and mutant yeast calmodulins by
 RT limited proteolysis and electrospray ionization mass spectrometry.";
 RL Protein Sci. 1:504-516(1992).
 RN [5]
 RP MUTAGENESIS.
 RX MEDLINE=91256315; PubMed=2044154;
 RA Geiser J.R., van Tuinen D., Brockerhoff S.E., Neff M.M., Davis T.N.;
 RT "Can calmodulin function without binding calcium?";
 RL Cell 65:949-959(1991).

[6]

RN STRUCTURE BY NMR OF 1-78.
 RX MEDLINE=96424985; PubMed=8827436;
 RA Ohki S.-Y., Miura K., Saito M., Nakashima K.-I., Maekawa H.,
 RA Yazawa M., Tsuda S., Hikichi K.;
 RT "Secondary structure and Ca(2+)-binding property of the N-terminal
 RT half domain of calmodulin from yeast *Saccharomyces cerevisiae* as
 RT studied by NMR.";
 RL J. Biochem. 119:1045-1055(1996).
 CC -|- FUNCTION: CALMODULIN MEDIATES THE CONTROL OF A LARGE NUMBER OF
 CC ENZYMES BY CA(++) AMONG THE ENZYMES TO BE STIMULATED BY THE
 CC CALMODULIN-CA(++) COMPLEX ARE A NUMBER OF PROTEIN KINASES AND
 CC PHOSPHATASES.
 CC -|- PTM: THE N-TERMINUS IS BLOCKED.
 CC -|- MISCELLANEOUS: THIS PROTEIN HAS THREE FUNCTIONAL CALCIUM-BINDING
 CC SITES.
 CC -|- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL: M14760; AAA34504.1; -
 DR EMBL: X78993; CAA55612.1; -
 DR EMBL: Z35978; CAA85064.1; -
 DR PIR: A25060; MCBY.
 DR HSSP: P02593; ICDM.
 DR SGD: S0000313; CMD1.
 DR InterPro: IPR002048; EF-hand.
 DR Pfam: PF00036; ehand; 4.
 DR ProDom: PD000012; EF-hand; 2.
 DR SMART: SM00054; EFh; 3.
 DR PROSITE: PS00018; EF_HAND; 3.
 KW Calcium-binding; Repeat.
 FT CA_BIND 21 32 EF-HAND 1.
 FT CA_BIND 57 68 EF-HAND 2.
 FT CA_BIND 94 105 EF-HAND 3.
 FT DOMAIN 129 140 ANCESTRAL CALCIUM SITE 4.
 FT MUTAGEN 21 21 D->A: HIGHLY REDUCED AFFINITY FOR CA++.
 FT MUTAGEN 32 32 D->V: HIGHLY REDUCED AFFINITY FOR CA++.
 FT MUTAGEN 57 57 D->A: HIGHLY REDUCED AFFINITY FOR CA++.
 FT MUTAGEN 68 68 E->V: HIGHLY REDUCED AFFINITY FOR CA++.
 FT MUTAGEN 94 94 D->A: HIGHLY REDUCED AFFINITY FOR CA++.
 FT MUTAGEN 105 105 E->V: HIGHLY REDUCED AFFINITY FOR CA++.
 SQ SEQUENCE 147 AA; 16135 MW; 819ED1AD5D9400D3 CRC64;

Query Match 14.7%; Score 59.5; DB 1; Length 147;
 Best Local Similarity 27.9%; Pred. No. 6.8;
 Matches 24; Conservative 11; Mismatches 30; Indels 21; Gaps 3;

QY 4 EHEVNLLVEEHRIGSKNADGKLSVKFG-----VLFQDDRCANLFEALV-----G 48
 Db 46 EAEVDNLNNEI-----DVDGNHQIEFSFLALMSQLKNSDSEQLLEAFKVFDRKNGDC 99
 QY 49 TLKAARRKRVITYAGELLQGVHDDV 74
 Db 100 LLSAAELKHVLTSGIKLTDAEVDDM 125

Search completed: January 17, 2003, 02:14:26
 Job time : 81 secs

```

RN SEQUENCE FROM N.A.
RP Porter M.B., Green M.L., Simmen R.C.M., Sharp D.C.;
RA "Cloning and sequencing of the equine pituitary gonadotropin-releasing
RT hormone receptor and expression of its mRNA following administration
RT of pulsatile or continuous GnRH.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS RECEPTOR MEDIATES ITS ACTION BY ASSOCIATION WITH
CC G PROTEINS THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC MESSENGER SYSTEM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF018072; AAC27099.2;
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm.1.1.
DR PRINTS: PR00237; GPCRHOPOPSN.
DR PROSITE: PS00237; G_PROTEIN_REC_F1_1; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 38 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 39 59 1 (POTENTIAL).
FT DOMAIN 60 84 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 85 105 2 (POTENTIAL).
FT DOMAIN 106 115 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 116 136 3 (POTENTIAL).
FT DOMAIN 137 157 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 158 178 4 (POTENTIAL).
FT DOMAIN 179 208 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 209 229 5 (POTENTIAL).
FT DOMAIN 230 271 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 272 292 6 (POTENTIAL).
FT DOMAIN 293 306 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 307 327 7 (POTENTIAL).
FT DOMAIN 328 338 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 18 18 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DISULFID 114 196 BY SIMILARITY.
SQ SEQUENCE 328 AA; 37648 MW; ADD7E25A41EACD0A CRC64;

Query Match 14.9%; Score 60.5; DB 1; Length 328;
Best Local Similarity 36.4%; Pred. No. 13;
Matches 20; Conservative 6; Mismatches 28; Indels 1; Gaps 1;

QY 13 EIHRLGSKNADGKLSVKGVLFGDDRCANLFEAL-VGTLKAARRKIKVTYVAGELL 66
Db 59 KLRHWTKRNGKRLSKMKVLLKHLIANLLEFLIWPDLGMNITVQVYVAGELL 113

RESULT 13
ARGB_BUCAI
ID ARGB_BUCAI STANDARD; PRT; 257 AA.
AC P57157;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acetylglutamate kinase (EC 2.7.2.8) (NAG kinase) (AGK) (N-acetyl-L-
DE glutamate 5-phosphotransferase).
GN ARGB OR BU049.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;

```

```

RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: ATP + N-acetyl-L-glutamate -> ADP + N-acetyl-L-
CC glutamate 5-phosphate.
CC -1- PATHWAY: Arginine biosynthesis; second step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE ACETYLGLUTAMATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001118; BAB12772.1;
DR InterPro: IPR001048; Aa_kinase.
DR Pfam: PF004662; AcgluKinase.
DR TIGRfam: TIGR00761; argB; 1.
KW Arginine biosynthesis; Transferase; Kinase; Complete proteome.
SQ SEQUENCE 257 AA; 27810 MW; 54EF26B46D6C064B CRC64;

Query Match 14.8%; Score 60; DB 1; Length 257;
Best Local Similarity 34.0%; Pred. No. 11;
Matches 18; Conservative 8; Mismatches 25; Indels 2; Gaps 1;

QY 26 LSVKFG--VLFQDDRCANLFEALVGTLKAARRKIKVTYVAGELLGVHDDVDI 76
Db 4 LVILGGVLLSESDAMKRLFEALVDYQFYKRHSVVIHGGRLIDNMNKL 56

RESULT 14
HELI_HSV6U
ID HELI_HSV6U STANDARD; PRT; 824 AA.
AC P52356;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable helicase.
GN U77 OR HDRF2.
OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
OX NCBI_TaxID=10370;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95027704; PubMed=7941342;
RA Nicholas J.;
RT "Nucleotide sequence analysis of a 21-kbp region of the genome of
RT human herpesvirus-6 containing homologues of human cytomegalovirus
RT major immediate-early and replication genes.";
RL Virology 204:738-750(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95266321; PubMed=7747482;
RA Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RA Martin M.E., Efstathiou S., Craxton M., Macaulay H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution.";
RL Virology 209:29-51(1995).
CC -1- FUNCTION: THIS PROTEIN MAY BE AN HELICASE AND IS REQUIRED FOR
CC REPLICATION OF VIRAL DNA.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL5,
CC EBV-1 57, EBV BBLF4, HCMV ULL105, AND VZV 55.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Serine/threonine-protein kinase Pim-3 (EC 2.7.1.37) (Pim-1).
GN PIM3 OR PIM1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A., AND PHOSPHORYLATION SITES.
RX MEDLINE=97256766; PubMed=9099695;
RA Palaty C.K., Kaimar G., Tai G., Oh S., Amankawa L., Affolter M.,
RA Aebersold R., Pelech S.L.;
RT "Identification of the autophosphorylation sites of the Xenopus
RL laevis Pim-1 proto-oncogene-encoded protein kinase.";
RL J. Biol. Chem. 272:10514-10521(1997).
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- PIM: Autophosphorylated.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC PIM SUBFAMILY.
CC -!- CAUTION: Was originally (Ref.1) called Pim-1 but seems to
CC represent the pim-3 isoform.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L29495; AAA85389.1;
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation.
FT DOMAIN 40 291 PROTEIN KINASE.
FT NP_BIND 46 54 ATP (BY SIMILARITY).
FT BINDING 69 69 ATP (BY SIMILARITY).
FT ACT_SITE 168 168 BY SIMILARITY.
FT MOD_RES 4 4 PHOSPHORYLATION (AUTO-).
FT MOD_RES 190 190 PHOSPHORYLATION (AUTO-).
FT MOD_RES 205 205 PHOSPHORYLATION (AUTO-).
SQ SEQUENCE 323 AA; 36964 MW; AB4DD61E7A99A38F CRC64;

Query Match 15.0%; Score 61; DB 1; Length 323;
Best Local Similarity 32.6%; Pred. No. 11;
Matches 29; Conservative 13; Mismatches 27; Indels 20; Gaps 6;

QY 2 NVEH-EVNL-----VEEIHRLGSKNADGKLSVFGVLFDQDRCANLFEALVGTILK 51
DB 1 NMEHLPVKILQPVKVDKEPFKEKVGQSVWASG-----GFGTVYSRIAD--GQPVAVKH 70

QY 52 AAKRKIVTYAGELLLQGVDDVDVILQ 80
DB 71 VAKER--VTEWG--TLNGVMVPLEIVLLK 95

RESULT 11
OPPF_HAEIN
ID OPFF_HAEIN STANDARD; PRT; 332 AA.
AC P45051;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Oligopeptide transport ATP-binding protein oppf.
GN OPFF OR H1120.

OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Spriggs T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR OLIGOPEPTIDES. PROBABLY RESPONSIBLE FOR ENERGY COUPLING TO THE
CC TRANSPORT SYSTEM (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (Potential).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32792; AAC22774.1;
DR TIGR; H1120;
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transportr.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transportr; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; Inner membrane; ATP-binding;
KW Peptide transport; Transport; Complete proteome.
FT NP_BIND 56 63 ATP (POTENTIAL).
SQ SEQUENCE 332 AA; 36760 MW; 45B6A3085431E3D9 CRC64;

Query Match 15.0%; Score 61; DB 1; Length 332;
Best Local Similarity 33.3%; Pred. No. 11;
Matches 22; Conservative 4; Mismatches 26; Indels 14; Gaps 1;

QY 1 MNVEHEVNLVVEIHRGLGSKNADGKLSVFGVLFDQDRCANLFEALVGTILKAARRKIYV 60
DB 1 MIVSNKKELLEVYNH-----LGVSFKIKNDKSLFFAKPQTLKAVKDVSKL 46

QY 61 YAGELL 66
DB 47 YAGETL 52

RESULT 12
GRHR_HORSE
ID GRHR_HORSE STANDARD; PRT; 328 AA.
AC Q18821;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Gonadotropin-releasing hormone receptor (GNRH-R).
GN GNRHR.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;

```



```

RX MEDLINE=92192451; PubMed=1532158;
RA Jones B.A., Fangman W.L.;
RT "Mitochondrial DNA maintenance in yeast requires a protein containing
RL a region related to the GTP-binding domain of dynamin.";
RL Genes Dev. 6:380-389(1992).
RN [2]
RX SEQUENCE OF 60-902 FROM N.A.
RX MEDLINE=93365024; PubMed=7916673;
RA Guan K., Farh L., Marshall T., Deschenes R.J.;
RT "Normal mitochondrial structure and genome maintenance in yeast
RL requires the dynamin-like product of the MGM1 gene.";
RL Curr. Genet. 24:141-148(1993).
RN [3]
RX SEQUENCE OF 1-805 FROM N.A.
RX STRAIN=5288C / FY1679;
RL MEDLINE=96437977; PubMed=8840505;
RA Gallison F., Dujon B.;
RT "Sequence and analysis of a 33 kb fragment from the right arm of
RL chromosome XV of the yeast Saccharomyces cerevisiae.";
RL Yeast 12:877-885(1996).
RN [4]
RX SEQUENCE OF 179-902 FROM N.A.
RA Hughes B., Pohl T.M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RX SEQUENCE OF 734-902 FROM N.A.
RX STRAIN=FL100;
RL MEDLINE=93296170; PubMed=8516295;
RA Lalo D., Carles C., Sentenac A., Thuriaux P.;
RT "Interactions between three common subunits of yeast RNA polymerases
RL I and III.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5524-5528(1993).
CC -I- FUNCTION: REQUIRED FOR MITOCHONDRIAL GENOME MAINTENANCE.
CC -I- SUBCELLULAR LOCATION: Mitochondrial.
CC -I- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62834; CAA44637.1; -
DR EMBL; L07419; -; NOT_ANNOTATED_CDS.
DR EMBL; X75119; CAA99426.1; -
DR EMBL; X92441; CAA63174.1; -
DR EMBL; L11274; AAB59316.1; -
DR EMBL; L75120; CAA99428.1; -
DR PIR; S33918; S33918.
DR PIR; S19068; S19068.
DR SGD; S0005737; MGM1.
DR InterPro; IPR001401; Dynamin.
DR Pfam; PF00350; dynamin; 1.
DR PRINTS; P00195; DYNAMIN.
DR SMART; SM00053; DYNC; 1.
DR PROSITE; PS00410; DYNAMIN; 1.
KW Motor protein; GTP-binding; Transit peptide; Mitochondrion.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 902 MGM1 PROTEIN.
FT DOMAIN 170 199 ASP-RICH (ACIDIC).
FT NP_BIND 238 245 GTP (POTENTIAL).
FT NP_BIND 338 342 GTP (POTENTIAL).
FT NP_BIND 406 409 GTP (POTENTIAL).
FT CONFLICT 150 150 G -> C (IN REF. 1 AND 2).
SQ SEQUENCE 902 AA; 101523 MW; 89B8C745182AA5E8 CRC64;

Query Match 15.4%; Score 62.5; DB 1; Length 902;
Best Local Similarity 25.6%; Pred No. 23;
Matches 23; Conservative 18; Mismatches 28; Indels 21; Gaps 4;

QY 4 EHEYNLLVEIHRLLGS-----KNADG--KLISVFGVLFODDRCANLFALVGTLLKAA--- 53

```

```

Db 717 QHSINLIKEELRQCSRYQATIKNAVSGKKLANVAGYLENE---SNLOKETIGMSKLLER 773
QY 54 -----KREKIVTYAGELLQGVHDDVD 75
Db 774 GSEAFIDKRCVKVLSFRLKMLKNKCHSTIE 803

RESULT 8
CY41_TRYBB
ID CY41_TRYBB STANDARD; PRT; 1242 AA.
AC Q99279;
DC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Receptor-type adenylate cyclase GRESAG 4.1 (EC 4.6.1.1) (ATP
DE pyrophosphate-lyase) (Adenylyl cyclase).
GN GRESAG 4.1.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EATRO 1125;
RX MEDLINE=91218809; PubMed=1982555;
RA Alexandre S., Paindavione P., Tebabi P., Pays A., Halleux S.,
RA Steinhert M., Pays E.;
RT "Differential expression of a family of putative adenylate/guanylate
RT cyclase genes in Trypanosoma brucei.";
RL Mol. Biochem. Parasitol. 43:279-288(1990).
CC -1- FUNCTION: COULD ACT AS A RECEPTOR FOR A UNKNOWN LIGAND.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52119; CAA36364.1; -.
CC InterPro; IPR001054; G_cyclase.
CC Pfam; PF00211; guanylate_cyc; 1.
CC SMART; SM00044; CYCC; 1.
CC DR PROSITE; PSS0125; GUANYLATE_CYCLASES_2; 1.
KW Lyase; camp synthetase; Transmembrane; Receptor; Glycoprotein.
FT DOMAIN 1 39
FT TRANSMEM 40 60
FT DOMAIN 61 862
FT TRANSMEM 863 883
FT DOMAIN 884 1242
FT CARBOHYD 116 116
FT CARBOHYD 289 289
FT CARBOHYD 318 318
FT CARBOHYD 338 338
FT CARBOHYD 401 401
FT CARBOHYD 534 534
FT CARBOHYD 563 563
FT CARBOHYD 603 603
FT CARBOHYD 702 702
FT CARBOHYD 741 741
FT CARBOHYD 818 818
SQ SEQUENCE 1242 AA; 138157 MW; 452F8B22FFC0A2E3 CRC64;

Query Match 15.3%; Score 62; DB 1; Length 1242;
Best Local Similarity 48.3%; Pred.No.38;
Matches 14; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 8 NLLVEEIHRLGSKNADGKLSVKFGVLFQD 36

```

[illegible]

```
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO M.JANNASCHII MJ0977.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67526; AAB98830.1; -
DR TIGR; MJ0831; -
DR InterPro; IPR003156; DHHA1.
DR InterPro; IPR001667; Ppsterase.
DR Pfam; PF01368; DHH; 1.
DR Pfam; PF02272; DHHA1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 432 AA; 48835 MW; 0EB95D188D2B41F3 CRC64;

Query Match 16.0%; Score 65; DB 1; Length 432;
Best Local Similarity 26.9%; Pred. No. 5.3;
Matches 18; Conservative 15; Mismatches 20; Indels 14; Gaps 3;

QY 3 VEHEVN---LLVEIHRGSKNADGKLSVKGVLFDQRCANL-----FEALVGTGK 51
DB IEHVRDAFLLEMLNVAVG---RNLFAVGIGICLEDDECICIGNQLWEYKKNLNLK 325
QY 52 AAKRRKI 58
DB 326 SVKLKLL 332

RESULT 3
VG17_BPP2A STANDARD; PRT; 174 AA.
AC P08389;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Early protein GP17.
GN 17.
OS Bacteriophage P2A.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC phi-29-like viruses.
OX NCBI_TaxID=10757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87031575; PubMed=3095189;
RA Paces V., Vicek C., Urbanek P., Hostomsky Z.;
RT "Nucleotide sequence of the right early region of Bacillus subtilis
RT phage P2A completes the 19366-bp sequence of P2A genome. Comparison
RT with the homologous sequence of phage phi 29.";
RL Gene 44:115-120(1986).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPLICATION OF THE
```

```
CC PHAGE DNA.
CC -1- SIMILARITY: BELONGS TO THE PODOVIRUSES GP17 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11813; AAA88499.1; -
DR PIR; F29004; WRBP70.
KW Early protein; DNA replication.
SQ SEQUENCE 174 AA; 20022 MW; 08FBCBF91D4B4224 CRC64;

Query Match 15.9%; Score 64.5; DB 1; Length 174;
Best Local Similarity 32.4%; Pred. No. 2.2;
Matches 22; Conservative 12; Mismatches 31; Indels 3; Gaps 2;

QY 9 LLVEIHRGSKNADGKLSVKGVLFDQRCANL-FEALVGTGKAAKRRKIYTAGELL 67
DB LLEVTVEELSGSDNDLDEIDSELYDEAASDVPHETIVKVFADK---SIVTFNGEK 119
QY 68 QGVHDDVD 75
DB 120 HYVNVVDV 127

RESULT 4
HV2I_HUMAN STANDARD; PRT; 146 AA.
AC P06331;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain V-II region ARH-77 precursor.
DE Ig heavy chain V-II region ARH-77 precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85205332; PubMed=3922855;
RA Kudo A., Ishihara T., Nishimura Y., Watanabe T.;
RT "A cloned human immunoglobulin heavy chain gene with a novel direct-
RT repeat sequence in 5' flanking region.";
RL Gene 33:181-189(1985).
DR PIR; A02101; G1HUH2.
DR HSP; P01825; 7FAB.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR Immunoglobulin V region; Signal.
DR SIGNAL 1 19
FT CHAIN 20 146 IG HEAVY CHAIN V-II REGION ARH-77.
FT DOMAIN 20 117 V SEGMENT.
FT DOMAIN 118 127 D SEGMENT.
FT DOMAIN 128 146 J SEGMENT.
FT DISULFID 42 115 BY SIMILARITY.
FT NON_TER 146 146
SQ SEQUENCE 146 AA; 16228 MW; 8D7FD52BB218171F CRC64;

Query Match 15.8%; Score 64; DB 1; Length 146;
Best Local Similarity 35.4%; Pred. No. 2.1;
Matches 23; Conservative 8; Mismatches 30; Indels 4; Gaps 2;

QY 11 VEETHRLGSKNADGKLSVKGVLFDQRCANL-FEALVGTGKAAKRRKIYTAGELL 70
DB IGEINHSGSTNY--KTSLSKRSVTISLDYTKNLFSLKLSVTAAD--TAVYICARGLRG 123
QY 71 HDDVD 75
```


Search completed: January 17, 2003, 02:17:36
Job time : 52 secs

Db 124 WNDVD 128

RESULT 11
H72336
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: H72336
R:Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.

Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: H72336
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-599 <ARN>
A:Cross-references: GB:AE001746; GB:AE000512; NID:g4981285; PIDN:AAD35852.1; PID:g498129
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0770

Query Match 15.8%; Score 64; DB 2; Length 599;
Best Local Similarity 36.5%; Pred. No. 36;
Matches 31; Conservative 9; Mismatches 37; Indels 8; Gaps 4;

QY 2 NVEHEVNLLVEITHRLGSKNADGKLSV---KFGVLFODDRCANLFEAL--VGTLLKAARR 56
Db 411 DLEEMVKLL-ERVKRLGFEVSASVSFVPEKHTPTFOFARQISPEEAEKIKTLKRAKRS 469
QY 57 KIVTYAGE--LLQGVHDDVDIVLL 79
Db 470 ARISYHDPMSLLEGVFSRGDRKLL 494

RESULT 12
F84784
probable glucosyl transferase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: F84784
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:1061197
A:Accession: F84784
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-496 <STO>
A:Cross-references: GB:AE002093; NID:g4415923; PIDN:AAD20154.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g36780
A:Map position: 2
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 15.6%; Score 63.5; DB 2; Length 496;
Best Local Similarity 32.1%; Pred. No. 33;
Matches 25; Conservative 11; Mismatches 27; Indels 15; Gaps 4;

QY 11 VEETHRLGSKNADGKLSVFGVLFODDRCANLFEALVGTLLKAAR-RRKIVTYAGELLQ 69
Db 418 VEEVKNKGEED-----KIGVLVDKEGYKAAVEELMGSDDAKERRRRVKELGELAHKA 470
QY 70 V-----HDDVDIVLLQD 81
Db 471 VEKGGSHSNITL-LLQD 487

RESULT 13

RESULT 9
WRBP70
early protein gp17 - phage PZA
C:Species: phage PZA
A:Note: host Bacillus subtilis
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 23-Jul-1999
C:Accession: F29004
R:Paces, V.; Vicek, C.; Urbanek, P.; Hostomsky, Z.
Gene 44, 115-120, 1986
A:Title: Nucleotide sequence of the right early region of Bacillus subtilis phage PZA
A:Reference number: A91551; MUID:87031575; PMID:3095189
A:Accession: F29004
A:Molecule type: DNA
A:Residues: 1-174 <PAC>
A:Cross-references: GB:M11813; GB:M13904; GB:M13905; NID:g216046; PIDN:AAA88499.1; PI
C:Genetics: 17
A:Gene: 17
C:Superfamily: phage PZA early protein gp17
C:Keywords: early protein

Query Match 15.9%; Score 64.5; DB 1; Length 174;
Best Local Similarity 32.4%; Pred. No. 7.9;
Matches 22; Conservative 12; Mismatches 31; Indels 3; Gaps 2;

QY 9 LLVEIHRLGSKNADGKLSVFGVLQDDRCANL-FEALVGTLLKAARKRKIVTYAGELL 67
DB 62 LLEVTVEELGEQNDLDDLEIDSELYDAESVPHEITVKVFEADK--SIVTNGEKLK 119
QY 68 QGVHDDVD 75
DB 120 HYVNVDDVD 127

RESULT 10
G1HUH2
Ig heavy chain precursor V-II region (ARH-77) - human
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 02-Sep-1997
C:Accession: A02101
R:Kudo, A.; Ishihara, T.; Nishimura, Y.; Watanabe, T.
Gene 33, 181-189, 1985
A:Title: A cloned human immunoglobulin heavy chain gene with a novel direct-repeat se
A:Reference number: A02101; MUID:85205332; PMID:392855
A:Accession: A02101
A:Molecule type: mRNA
A:Residues: 1-146 <KUD>
A:Note: the sequence was determined from the differentiated gene
A:Note: the authors translated the codon GGG for residue 17 as Arg
C:Genetics:
A:Gene: GDB:IGHV@
A:Cross-references: GDB:128528; OMIM:147070
A:Map position: 14q32.33-14q32.33
A:Introns: 16/3
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-146/Product: Ig heavy chain V-II region (ARH-77) #status predicted <MAT>
F:20-117/Region: V segment
F:35-117/Domain: immunoglobulin homology <IMM>
F:118-127/Region: D segment
F:128-146/Region: J segment
F:42-115/Disulfide bonds: #status predicted

Query Match 15.8%; Score 64; DB 1; Length 146;
Best Local Similarity 35.4%; Pred. No. 7.4;
Matches 23; Conservative 8; Mismatches 30; Indels 4; Gaps 2;

QY 11 VEEIHRLGSKNADGKLSVFGVLQDDRCANLFEALVGTLLKAARKRKIVTYAGELL 70
DB 68 ICEIHSGSTNY--KTSLSKRSVITSLDTSKNLPSLKLSSVTAAD--TAVYTCARGLLRGG 123
QY 71 HDDVD 75

Query Match 23.8%; Score 96.5; DB 2; Length 162;
Best Local Similarity 32.9%; Pred. No. 0.002;
Matches 26; Conservative 16; Mismatches 34; Indels 3; Gaps 2;

QY 2 NVEHEVNLVEEIHRLGSKNADG-KLSVKFGVLQDDRCANLFEALVGTTLKAARRKIVT 60
Db 66 HVHREMLTLCWEVDYKQKEGSPITRTFGRLF--TIYVNISDKVVGVTLLRARRKKMD 123
QY 61 YAGELLQGVHDDVDIVLL 79
Db 124 FEGEMLFQKRDDHVIILL 142

RESULT 3
T24421
hypothetical protein T04A8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24421
R:Palmer, S.
submitted to the EMBL Data Library, August 1994
A:Reference number: Z19889
A:Accession: T24421
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-289 <MIL>
A:Cross-references: EMBL:Z35663; PIDN:CAA84722.1; GSPDB:GN00021; CESP:T04A8.4
A:Experimental source: clone T04A8
C:Genetics:
A:Gene: CESP:T04A8.4
A:Map position: 3
A:Introns: 26/2; 58/3; 102/3; 191/2; 223/3

Query Match 20.1%; Score 81.5; DB 2; Length 289;
Best Local Similarity 33.7%; Pred. No. 0.18;
Matches 29; Conservative 14; Mismatches 26; Indels 17; Gaps 4;

QY 1 MNVEHEVNLVEEIHRLGSKNADGK---LSVKFGVLQDDRCANLFEALVGTTLKAA 53
Db 61 VHCREILFLCETI-----DSNADGEEPHKYVFKGLF-----NIYFYSKLVGMILRA 110

QY 54 KRRKIVTYAGELLQGVHDDVDIVLL 79
Db 111 RYGLVHFEGEMLYQRODDEKIIML 136

RESULT 4
E90541
hypothetical protein MYPV_2370 [Imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: E90541
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Sanson, D.; Gallisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: E90541
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-318 <KOR>
A:Cross-references: GB:AL445566; PID:gl4089650; PIDN:CAC13410.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV_2370
A:Genetic code: SGC3
C:Superfamily: phosphate acetyltransferase

Query Match 17.0%; Score 69; DB 2; Length 318;
Best Local Similarity 36.2%; Pred. No. 4.9;
Matches 29; Conservative 6; Mismatches 25; Indels 20; Gaps 4;

QY 6 EVNLLVEEIHRLGSKNADGKLSVKFGVLQDDRCANLFEALVGTTLKAAK 54

Db 44 EYSLIVE-----SKSDYKGFNFVLLDQDQKQYSEFCODLFESRKGRKDSLESVKALK 96
QY 55 RRRKIVTYAGELLQGVHDDV 74
Db 97 TRPF--YAMMLLKGFFDGV 114

RESULT 5
F81343
probable integral membrane zinc-metalloprotease Cj0723c [Imported] - Campylobacter jejuni
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: F81343
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chlil
Nature 403, 663-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: F81343
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-395 <PAR>
A:Cross-references: GB:AL139076; GB:AL111168; NID:g6968128; PIDN:CAB72997.1; PID:g696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0723c

Query Match 16.9%; Score 68.5; DB 2; Length 395;
Best Local Similarity 26.2%; Pred. No. 7.1;
Matches 21; Conservative 19; Mismatches 29; Indels 11; Gaps 3;

QY 2 NVEHEVNLVEE-----IHLGSKNADGKLSVKFGVLQDDRCANLFEALVGTTLKAAK 54
Db 200 NLLKISSLMKQCGFANGVYVIDASKRDLNAYEGGLFKSRVV-LFDTL---LKALN 255
QY 55 RRRKIVTYAGELLQGVHDDV 74
Db 256 ERELLAVLGHELGHFVHKDI 275

RESULT 6
T04871
hypothetical protein F28A21.200 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
C:Accession: T04871
R:Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15387
A:Accession: T04871
A:Molecule type: DNA
A:Residues: 1-530 <BEV>
A:Cross-references: EMBL:AL035526
A:Experimental source: cultivar Columbia; BAC clone F28A21
C:Genetics:
A:Map position: 4
A:Introns: 193/2; 239/1; 456/3
A:Note: F28A21.200
C:Superfamily: natural resistance-associated macrophage protein 1

Query Match 16.3%; Score 66; DB 2; Length 530;
Best Local Similarity 29.5%; Pred. No. 19;
Matches 23; Conservative 11; Mismatches 36; Indels 8; Gaps 3;

QY 3 VEHEVNLVEEIHRLGSKNADGKLSVKFGVLQDDRCANLFEALVGTTLKAAK 55
Db 303 VSEMINLFVTAFAKGFYGTQKQADSIGLVNAGVYLQEKYGGVFPILYIWIGIGLLAAGS 362

QY 56 RKIV-TYAGELLQGVHDDV 72
Db 363 STITGTAGQFTMEGFLD 380

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	286	70.4	161	2	T04981	hypothetical prote
2	96.5	23.8	162	2	T21860	hypothetical prote
3	81.5	20.1	289	2	T24421	hypothetical prote
4	69	17.0	318	2	E90541	hypothetical prote
5	68.5	16.9	395	2	F81343	probable integral
6	66	16.3	530	2	T04871	hypothetical prote
7	66	16.3	1404	2	E36788	hypothetical prote
8	65	16.0	432	2	G64403	hypothetical prote
9	64.5	15.9	174	1	WRBP70	early protein gp17
10	64	15.8	146	1	G1H0H2	ig heavy chain pre
11	64	15.8	599	2	T27336	conserved hypothet
12	63.5	15.6	496	2	F84784	probable glucosyl
13	63	15.5	514	2	T29652	inward rectifier p
14	63	15.5	811	2	S76690	hypothetical prote
15	63	15.5	2535	2	T04824	hypothetical prote
16	62.5	15.4	491	2	C84784	probable glucosyl
17	62.5	15.4	496	2	D84784	probable glucosyl
18	62.5	15.4	565	2	G64532	methyl-accepting c
19	62.5	15.4	902	2	S39318	dynamlin-like prote
20	62	15.3	268	2	AE2515	hypothetical prote
21	62	15.3	1442	2	S14201	probable adenylate
22	61.5	15.1	637	1	WQEC2M	phosphotransferase
23	61.5	15.1	988	2	AE3697	glutamate-ammonia-
24	61.5	15.1	1010	2	E97479	hypothetical prote
25	61	15.0	217	2	AH0561	potential acRAB op
26	61	15.0	332	2	B64184	oligopeptide trans
27	61	15.0	556	2	A81931	probable adhesin N
28	61	15.0	635	2	T15904	hypothetical prote
29	60.5	14.9	554	2	F71975	methyl-accepting c

Fri Jan 17 09:17:31 2003

us-09-648-310-2.ra1

Page 6

Search completed: January 17, 2003, 02:18:24
Job time : 37 secs

FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4164
LENGTH: 128
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4164

Query Match 13.4%; Score 54.5; DB 4; Length 128;
Best Local Similarity 30.6%; Pred. No. 8.7;
Matches 26; Conservative 13; Mismatches 35; Indels 11; Gaps 3;
QY 4 EHEVNLVEEIHRLGSK-----KNADGKLSVKFGVLFQDDRCANLFEALVGTLLKAARR 56
Db 22 QYELEKLLKENEELKAEKALSQMKNETRSMNLNESGLENFDDQIVNI--LVNT-DAEKTR 77
QY 57 KIVTYAGELLQGVHDDVDVILQD 81
Db 78 KNVESPTNLLNQMKSVNKEALRQD 102

RESULT 13
US-09-058-489-18
Sequence 18, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WHI97-08pA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 18
LENGTH: 660
TYPE: PRT
ORGANISM: Human
US-09-058-489-18

Query Match 13.4%; Score 54.5; DB 3; Length 660;
Best Local Similarity 27.1%; Pred. No. 69;
Matches 16; Conservative 12; Mismatches 24; Indels 7; Gaps 2;
QY 2 NVEHEVNL-----LVEEIHRLGSKNADGKLSVKFGVLFQDDRCANLFEALVGTLLKAARR 55
Db 507 NVRHVINFDPSPDIEYVHRIGTRGVNGL--ATSFNEKNMNITKDLLDLVLEAKQ 563

RESULT 14
US-09-058-489-91
Sequence 91, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WHI97-08pA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877

EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 91
LENGTH: 660
TYPE: PRT
ORGANISM: Human
US-09-058-489-91

Query Match 13.4%; Score 54.5; DB 3; Length 660;
Best Local Similarity 27.1%; Pred. No. 69;
Matches 16; Conservative 12; Mismatches 24; Indels 7; Gaps 2;
QY 2 NVEHEVNL-----LVEEIHRLGSKNADGKLSVKFGVLFQDDRCANLFEALVGTLLKAARR 55
Db 507 NVRHVINFDPSPDIEYVHRIGTRGVNGL--ATSFNEKNMNITKDLLDLVLEAKQ 563

RESULT 15
US-08-952-127-12
Sequence 12, Application US/08952127
Patent No. 6211336
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
APPLICANT: Tagle, Danilo A.
APPLICANT: Collins, Francis S.
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6211336thwestern Hwy., Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,127
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: 2290, 00029
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3066 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Mus musculus
US-08-952-127-12

Query Match 13.4%; Score 54.5; DB 4; Length 3066;
Best Local Similarity 26.6%; Pred. No. 4.7e+02;
Matches 25; Conservative 16; Mismatches 28; Indels 25; Gaps 5;
QY 8 NLLVEEIHRLGSK-----NADGKLSVKFGVL-FQDDRCANLFEALVGTLLKAA----- 53
Db 317 DLLVNEISHIGSRKYSGRNIANKENLIDLMDADICQLFDADTRSVESQSVVTQRES 376
QY 54 -----KRRKIVTYAGELL---LQGVHDDVDIV 77
Db 377 TDYSVPCKRRKI-DVGWEVINKDYLKQSQSDFLV 409

```

US-08-465-250-2
Query Match      13.8%   Score 56; DB 4; Length 2206;
Best Local Similarity 36.2%; Pred.No. 2e+02;
Matches 17; Conservative    6; Mismatches    8; Indels    16; Gaps    2;

Oy  47 VGTLLKAAK-----RRKIVTYA---GELLILQGVHDDVDIV 77
     | :|::|||          :| |||         :| ::|||::| :|
Db  1553 VPTIRAAKVQSGFDYAVAMAKRNIVTATTSGKEFTMLGVHDNVAIL 1599

RESULT 11
US-08-203-905B-14
; Sequence 14, Application US/08203905B
; Patent No. 5646249
; GENERAL INFORMATION:
; APPLICANT: KAYE, FEDERIC J.
; APPLICANT: OTTERSON, GREGORY A.
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A
; TITLE OF INVENTION: NOVEL CHAPERONE PROTEIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CA
; COUNTRY: USA
; ZIP: 92660

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,905B
; FILING DATE: February 28, 1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: KIRKPATRICK, ANITA M.
; REGISTRATION NUMBER: 32,617
; REFERENCE/DOCKET NUMBER: NIH089.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-203-905B-14

Query Match 13.7%; Score 55.5; DB 1; Length 472;
Best Local Similarity 27.0%; Pred. No. 33;
Matches 27; Conservative 12; Mismatches 30; Indels 31; Gaps 3;

QY 7 VLLVEEIHRLGSKNADGKL-----SVKFGVLFQDDRCAN 41
| | | | | | | | | | | | | | | | | | | | |
Ddb 310 VLLTVEENDSQPNADSKLPDQTPGQHVNVRFGLSDTSAKSQVLFETEVSRK 369
| | | | | | | | | | | | | | | | | | | | |

QY 42 LFEALVTGLAAKRRIKVIYAGELLIQGVHD--DVDIVLL 79
| | | | | | | | | | | | | | | | | | | | |
Ddb 370 LF---NTLNEDLFQILVPIQVLKEGLDKTEIDVVL 405
| | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-09-134-001C-4164
; Sequence 4164, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

```

;
; LENGTH: 2052 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
US-09-045-201A-2

Query Match 13.8%; Score 56; DB 3; Length 2052;
Best Local Similarity 26.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 15; Mismatches 23; Indels 16; Gaps 4;

QY 14 IHRIGSKNADGKLSVKFGV-----LFQDDRCANFEALVGLTKAAKRKIYTYAGELLQ 68
Db 1870 VIRIGYKNSQNTTEKKLLDVNENLFYGRKMAQVFD-LKGSLL---RRNRVKTDTGK---- 1921

QY 69 GVHDDVDIVLQD 81
Db 1922 ---ESCDVLLDE 1931

RESULT 7
US-09-619-062-2
; Sequence 2, Application US/09619062
; Patent No. 6406875
; GENERAL INFORMATION:
; APPLICANT: Shisheva, Assia
; TITLE OF INVENTION: No. 6406875el Mammalian Putative
; TITLE OF INVENTION: Phosphatidylinositol-4-Phosphate-5-Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/619,062
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/045,201
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Smith, DeAnn F
REFERENCE/DOCKET NUMBER: 4981-098431
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 641-1600
TELEFAX: (248) 641-0270
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
US-09-619-062-2

Query Match 13.8%; Score 56; DB 4; Length 2052;
Best Local Similarity 26.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 15; Mismatches 23; Indels 16; Gaps 4;

QY 14 IHRIGSKNADGKLSVKFGV-----LFQDDRCANFEALVGLTKAAKRKIYTYAGELLQ 68
Db 1870 VIRIGYKNSQNTTEKKLLDVNENLFYGRKMAQVFD-LKGSLL---RRNRVKTDTGK---- 1921

QY 69 GVHDDVDIVLQD 81
Db 1922 ---ESCDVLLDE 1931

RESULT 8
US-07-852-260-2
; Sequence 2, Application US/07852260
; Patent No. 5525715
; GENERAL INFORMATION:
; APPLICANT: Racanietello, Vincent
; APPLICANT: Tatem, Joanne M.
; APPLICANT: Weeks-Levy, Carolyn L.
; TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES FROM
; TITLE OF INVENTION: CDNA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/852,260
FILING DATE: 19920619
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 36607-B-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2206 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-852-260-2

Query Match 13.8%; Score 56; DB 1; Length 2206;
Best Local Similarity 36.2%; Pred. No. 2e+02;
Matches 17; Conservative 6; Mismatches 8; Indels 16; Gaps 2;

QY 47 VGTLLKAAK-----RRKIYTYA---GELLQGVHDDVDIV 77
Db 1553 VPIRAAKVQPGFVAVAMAKRNIYATTSKGEFTMLGVHDNAIL 1599

RESULT 9
US-08-461-503-2
; Sequence 2, Application US/08461503
; Patent No. 5834302
; GENERAL INFORMATION:
; APPLICANT: Racanietello, Vincent
; APPLICANT: Tatem, Joanne M.
; APPLICANT: Weeks-Levy, Carolyn L.
; TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES
; TITLE OF INVENTION: FROM CDNA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```
Query Match          15.0%; Score 61; DB 4; Length 323;
Best Local Similarity 32.6%; Pred. No. 3.8;
Matches 29; Conservative 13; Mismatches 27; Indels 20; Gaps 6;

QY 2 NVEH-EVNL-----VEEHLRLGSKNADGKLSVKFGVLFQDDRCANLFEALVGTAKR 51
Db 17 NMEHLVPKILQPVKDKPEFKYQGVSVASG-----GFGTVYSDSRIAD--GQPVAVKH 70

QY 52 AAKRRKIVTYAGELLQGVHDDVDIVLLQ 80
Db 71 VAKER--VTEWG--TLNGVMVPLEIVLK 95

RESULT 3
US-09-058-489-15
; Sequence 15, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WH197-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-15

Query Match          14.4%; Score 58.5; DB 3; Length 662;
Best Local Similarity 28.8%; Pred. No. 20;
Matches 17; Conservative 12; Mismatches 23; Indels 7; Gaps 2;

QY 2 NVEH-EVNL-----LVEEHLRLGSKNADGKLSVKFGVLFQDDRCANLFEALVGTAKR 55
Db 509 NVKHVINFDLPDSIEEYVHRIGTRGVNGL--ATSPFNERNINITKDLLLVEAKQ 565

RESULT 4
US-09-058-489-16
; Sequence 16, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WH197-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-16

Query Match          14.4%; Score 58.5; DB 3; Length 662;
Best Local Similarity 28.8%; Pred. No. 20;
Matches 17; Conservative 12; Mismatches 23; Indels 7; Gaps 2;
```

```
QY 2 NVEH-EVNL-----LVEEHLRLGSKNADGKLSVKFGVLFQDDRCANLFEALVGTAKR 55
Db 509 NVKHVINFDLPDSIEEYVHRIGTRGVNGL--ATSPFNERNINITKDLLLVEAKQ 565

RESULT 5
US-09-134-001C-2873
; Sequence 2873, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCI
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2873
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2873

Query Match          13.8%; Score 56; DB 4; Length 208;
Best Local Similarity 28.3%; Pred. No. 10;
Matches 22; Conservative 14; Mismatches 22; Indels 20; Gaps 4;

QY 10 LVEEHLRLGSKNADGKLSVKFGVLFQDDRCANLFEALVGTAKRRIKIVYAG 63
Db 29 MVRRCHSI--REARRLNVASSAVNRQILKLEDEMGATLFDRLPGLR-----VTAAG 79

QY 64 ELLQGVHDDVDIVLLQD 81
Db 80 EILTRHI-----TILLQD 92

RESULT 6
US-09-045-201A-2
; Sequence 2, Application US/09045201A
; Patent No. 6110718
; GENERAL INFORMATION:
; APPLICANT: Shisheva, Asia
; TITLE OF INVENTION: No. 6110718el Mammalian Putative
; TITLE OF INVENTION: Phosphatidylinositol-4-Phosphate-5-Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,201A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F
; REFERENCE/DOCKET NUMBER: 4981-098431
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 641-1600
; TELEFAX: (248) 641-0270
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 17, 2003, 02:04:14 ; Search time 36 Seconds
(without alignments)
66.202 Million cell updates/sec

Title: US-09-648-310-2

Perfect score: 406
Sequence: 1 MNVEHEVNLLVEEHRGSK.....AGELLQGVHDDVDIVLLQD 81

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	61	15.0	323	4	US-09-237-543-6
2	61	15.0	323	4	US-09-644-450-6
3	58.5	14.4	662	3	US-09-058-489-15
4	58.5	14.4	662	3	US-09-058-489-16
5	56	13.8	208	4	US-09-134-001C-2873
6	56	13.8	2052	3	US-09-045-201A-2
7	56	13.8	2052	4	US-09-619-062-2
8	56	13.8	2206	1	US-07-852-260-2
9	56	13.8	2206	2	US-08-461-503-2
10	56	13.8	2206	2	US-08-465-250-2
11	55.5	13.7	472	1	US-08-203-905B-14
12	54.5	13.4	128	4	US-09-134-001C-4164
13	54.5	13.4	660	3	US-09-058-489-18
14	54.5	13.4	660	3	US-09-058-489-91
15	54.5	13.4	3066	4	US-08-952-127-12
16	54	13.3	129	4	US-08-981-527A-11
17	54	13.3	864	2	US-08-209-521-29
18	53.5	13.2	402	4	US-09-347-801-16
19	53.5	13.2	466	4	US-08-914-375C-60
20	53.5	13.2	966	1	US-08-571-758-2
21	53.5	13.2	966	1	US-08-909-984A-2
22	53.5	13.2	966	1	US-08-909-983-2
23	53	13.1	289	4	US-09-071-035-72
24	53	13.1	312	4	US-09-071-035-70
25	53	13.1	367	2	US-08-530-230-24
26	52.5	12.9	138	4	US-09-134-001C-2982
27	52.5	12.9	689	1	US-07-766-351-5

```

28 52.5 12.9 689 1 US-08-059-032-5 Sequence 5, Appli
29 52.5 12.9 689 5 PCT-US91-07290-5 Sequence 5, Appli
30 52.5 12.9 739 2 US-08-836-943-2 Sequence 2, Appli
31 52 12.8 489 1 US-08-095-726-8 Sequence 8, Appli
32 52 12.8 489 1 US-08-095-726-10 Sequence 10, Appli
33 52 12.8 489 1 US-08-096-043-8 Sequence 8, Appli
34 52 12.8 489 1 US-08-096-043-10 Sequence 10, Appli
35 52 12.8 489 1 US-08-096-623A-8 Sequence 8, Appli
36 52 12.8 489 1 US-08-096-623A-10 Sequence 10, Appli
37 52 12.8 510 4 US-09-291-922-22 Sequence 22, Appli
38 51.5 12.7 155 2 US-08-844-057-4 Sequence 4, Appli
39 51.5 12.7 155 4 US-09-006-730-4 Sequence 4, Appli
40 51.5 12.7 315 4 US-09-134-001C-4771 Sequence 4771, Ap
41 51.5 12.7 425 4 US-09-134-001C-5619 Sequence 5619, Ap
42 51.5 12.7 872 2 US-08-844-057-2 Sequence 2, Appli
43 51.5 12.7 872 4 US-09-006-730-2 Sequence 2, Appli
44 51 12.6 946 3 US-08-560-005-4 Sequence 4, Appli
45 51 12.6 946 4 US-09-418-540-4 Sequence 4, Appli

```

ALIGNMENTS

```

RESULT 1
US-09-237-543-6
; Sequence 6, Application US/09237543A
; Patent No. 6143540
; GENERAL INFORMATION:
; APPLICANT: Kapeller, Rosana
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HKID-1-RELATED PROTEIN FAMILY
; FILE REFERENCE: 035800/175631
; CURRENT APPLICATION NUMBER: US/09/237,543A
; CURRENT FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Xenopus laevis
US-09-237-543-6

```

```

Query Match 15.0%; Score 61; DB 4; Length 323;
Best Local Similarity 32.6%; Pred. No. 3.8;
Matches 29; Conservative 13; Mismatches 27; Indels 20; Gaps 6;

```

```

QY 2 NVEH-EVNLL-----VEEHRGSKNADGKLSVKGVLFODDRCANLFEALVGLK 51
Db 17 NMEHLPVKILQPKVKDKPEFKVQGVSVASG-----GFGTVYSDSRIAD--GQPVAVKH 70

```

```

QY 52 AAKRRKIVYAGELLQGVHDDVDIVLLQ 80
Db 71 VAKER--VTWEG--TLNGVMVPLEIVLLK 95

```

```

RESULT 2
US-09-644-450-6
; Sequence 6, Application US/09644450
; Patent No. 6363791
; GENERAL INFORMATION:
; APPLICANT: Kapeller, Rosana
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HKID-1-RELATED PROTEIN FAMILY
; FILE REFERENCE: 035800/175631
; CURRENT APPLICATION NUMBER: US/09/644,450
; CURRENT FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Xenopus laevis
US-09-644-450-6

```

```

; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2516
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-2

Query Match 13.4%; Score 54.5; DB 10; Length 2516;
Best Local Similarity 23.9%; Pred. No. 3.2e+02;
Matches 26; Conservative 15; Mismatches 33; Indels 35; Gaps 4;

QY 1 MNYEHEVNL-----LVEEIHL-----GSKN-----ADGKLSVK 29
Db 2329 LEVERTVSLAEVYAGLPKNGPFLAQEIDKLVSQSGSAGSGNNLAFAGTDTKTSLQ 2388
;
QY 30 FGVLFDQDRCANLFEALVGLTKAAKRRKIITYAGELLQGVHDDVDIVL 78
Db 2389 ASVSFADLKIREDPASLGKIRRIKQISVTLPA-----LLGPYQDVQAIL 2433

RESULT 15
US-09-788-657-16
; Sequence 16, Application US/09788657
; Patent No. US20020123149A1
; GENERAL INFORMATION:
; APPLICANT: Nicolalides, Nicholas
; APPLICANT: Sass, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/09/788,657
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-788-657-16

Query Match 13.3%; Score 54; DB 10; Length 859;
Best Local Similarity 29.2%; Pred. No. 94;
Matches 14; Conservative 12; Mismatches 14; Indels 8; Gaps 2;

QY 26 LSVKFGVLFQDRCANLF-----EALVGT-LKAAKRRKIITYAGEL 65
Db 787 LSDSPGVCMRPSRVROMFASCRACKSVNIGTALNASEMKKLITHGEM 834

Search completed: January 17, 2003, 02:19:49
Job time : 73 secs

```

Search completed: January 17, 2003, 02:19:49
Job time : 73 secs

```

Query Match      13.5%; Score 55; DB 10; Length 1214;
Best Local Similarity 28.3%; Pred. No. 1.1e+02;
Matches 17; Conservative 12; Mismatches 17; Indels 14; Gaps 2;

yy      25 KLSVKGVLFQDDRCANLFEALVGTLLKAKRKRTVYAGELLQGVHDDV---DIVLQD 81
bb      91 RINAQGVLFQEDQIILDWFVQCLALKLKHVDKRI-----LHRDKSQNIFLTKD 139

RESULT 10

```

RESULT 12
S-09-115-150-4
Sequence 4, Application US/09115150

```

Db      447 GN-----GTLTRTKKSFWMYKKVIASNGE 471          |||   ::   :|||   ||
RESULT 7
US-09-815-242-10214
; Sequence 10214, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10214
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10214
Query Match           13.8%; Score 56; DB 10; Length 506;
Best Local Similarity 32.5%; Pred. No. 27;
Matches 26; Conservative 9; Mismatches 27; Indels 18; Gaps 5;

QY    7 VNLIV--EETHRLGSKNADGK---LSWKEGVLFODRCANIFEALVGLKAARRKVITY 61
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    32 VNLRPSPHSTIALMGANGAGCKSTLLKCLFGI-YOKDSGTILFQGKEIDFHSAK 83
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY    62 AGELLLOG---VHDDVDIVL 78
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    84 --EALENGISMVHQELNLVL 101
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 8
US-09-815-242-10047
; Sequence 10047, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A

```

```

; APPLICANT: Laura A. Rudolph-Owen
; APPLICANT: Kyle MacBeth
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HKID-1-RELATED PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 35800/238856
; CURRENT APPLICATION NUMBER: US/09/971,791
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 09/644,450
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/237,543
; PRIOR FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Xenopus laevis
US-09-971-791-6

Query Match      15.08; Score 61; DB 10; Length 323;
Best Local Similarity 32.6%; Pred. No. 3.5;
Matches 29; Conservative 13; Mismatches 27; Indels 20; Gaps 6;

Qy  2 NVEH-EVNL-----VEIHRIGSKNADGKLSVFGVLFQDDRCANLFEALVGTILK 51
Db  17 NMEHLPVKILQPVKDKPEFKVQGVSVASG-----CFGIVSDRIAD--GQPVAVKH 70

Qy  52 AAKRRKIVTYAGELLQGVHDDVDIVLQ 80
Db  71 VAKER--VTENG--TLNGVMVPLEIVLK 95

RESULT 3
US-09-815-242-5045
; Sequence 5045, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5045
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae

Query Match      14.08; Score 57; DB 10; Length 506;
Best Local Similarity 32.5%; Pred. No. 20;
Matches 26; Conservative 10; Mismatches 26; Indels 18; Gaps 5;

Qy  7 VNLIV--EETHRLGSKNADGK---LSVKFGLVQDDRCANLFEALVGTILKAARRKIVTY 61
Db  32 VNLKVRPHSIHALMGENGAGKSTLLKCLFGI-YOKDSGSLFQGKEIDFHSK----- 83

Qy  62 AGELLQG---VHDDVDIVL 78
Db  84 --EALENGISMVHQELNLVL 101

RESULT 4
US-09-815-242-11757
; Sequence 11757, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11757
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae

Query Match      14.08; Score 57; DB 10; Length 506;
Best Local Similarity 32.5%; Pred. No. 20;
Matches 26; Conservative 10; Mismatches 26; Indels 18; Gaps 5;

Qy  7 VNLIV--EETHRLGSKNADGK---LSVKFGLVQDDRCANLFEALVGTILKAARRKIVTY 61
Db  32 VNLKVRPHSIHALMGENGAGKSTLLKCLFGI-YOKDSGSLFQGKEIDFHSK----- 83

Qy  62 AGELLQG---VHDDVDIVL 78
Db  84 --EALENGISMVHQELNLVL 101

RESULT 5
US-09-756-526A-2
; Sequence 2, Application US/09756526A
; Patent No. US20020038005A1
; GENERAL INFORMATION:
; APPLICANT: Jana, Wojciechowska
; APPLICANT: Evgeny, Lewitin
; APPLICANT: Ludmila, Revina
; APPLICANT: Igor, Zalunin
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 17, 2003, 02:06:44 ; Search time 71 Seconds
(without alignments)
22.678 Million cell updates/sec

Title: US-09-648-310-2
Perfect score: 406
Sequence: 1 MNVEHEVNLVEIHRGSK.....AGELLQGVHDDVDIVLLQD 81

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications-AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	390	96.1	142	10	US-09-925-300-1485
2	61	15.0	323	10	US-09-971-791-6
3	57	14.0	506	10	US-09-815-242-5045
4	57	14.0	506	10	US-09-815-242-11757
5	57	14.0	1163	10	US-09-756-526A-2
6	56.5	13.9	474	10	US-09-815-242-10270
7	56	13.8	506	10	US-09-815-242-10214
8	55.5	13.7	1160	10	US-09-815-242-10047
9	55	13.5	1214	10	US-09-783-320-4
10	54.5	13.4	239	10	US-09-767-041-10
11	54.5	13.4	309	10	US-09-825-882-12
12	54.5	13.4	549	10	US-09-115-150-4
13	54.5	13.4	639	10	US-09-925-301-1158
14	54.5	13.4	2516	10	US-09-817-514A-2
15	54	13.3	859	10	US-09-788-657-16
16	53.5	13.2	162	10	US-09-864-761-43711
17	53.5	13.2	394	9	US-09-738-626-3505
18	53.5	13.2	402	10	US-09-854-731-16
19	53	13.1	193	10	US-09-815-242-11663

20	53	13.1	367	10	US-09-919-497-60	Sequence 60, Appl
21	53	13.1	506	10	US-09-864-761-46512	Sequence 46512, A
22	52.5	12.9	476	10	US-09-779-307-16	Sequence 16, Appl
23	52.5	12.9	477	10	US-09-779-307-4	Sequence 4, Appli
24	52.5	12.9	608	9	US-09-738-626-3609	Sequence 3609, Ap
25	52.5	12.9	689	9	US-10-113-085-2	Sequence 2, Appli
26	52.5	12.9	739	9	US-09-738-626-6988	Sequence 6988, Ap
27	52.5	12.9	771	9	US-09-712-363-157	Sequence 157, App
28	52.5	12.9	872	10	US-09-815-242-13594	Sequence 13594, A
29	52.5	12.9	876	10	US-09-815-242-12623	Sequence 12623, A
30	52	12.8	122	10	US-09-815-242-4935	Sequence 4935, Ap
31	52	12.8	122	10	US-09-815-242-10525	Sequence 10525, A
32	52	12.8	271	10	US-09-815-242-11624	Sequence 11624, A
33	52	12.8	510	9	US-10-051-902-22	Sequence 22, Appl
34	52	12.8	510	9	US-10-051-909-22	Sequence 22, Appl
35	51.5	12.7	272	9	US-09-738-626-5598	Sequence 5598, Ap
36	51.5	12.7	296	10	US-09-843-164-4	Sequence 4, Appli
37	51.5	12.7	426	10	US-09-815-242-5383	Sequence 5383, Ap
38	51.5	12.7	429	10	US-09-815-242-12502	Sequence 12502, A
39	51.5	12.7	751	10	US-09-864-761-38419	Sequence 38419, A
40	51.5	12.7	850	10	US-09-888-615-108	Sequence 108, App
41	51.5	12.7	1175	10	US-09-771-161A-224	Sequence 224, App
42	51.5	12.7	1175	10	US-09-771-161A-225	Sequence 225, App
43	51.5	12.7	1175	10	US-09-771-161A-226	Sequence 226, App
44	51.5	12.7	2025	10	US-09-815-242-5703	Sequence 5703, Ap
45	51.5	12.7	3158	10	US-09-815-242-12611	Sequence 12611, A

ALIGNMENTS

RESULT 1
US-09-925-300-1485
; Sequence 1485, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1485
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1485

Query Match 96.1%; Score 390; DB 10; Length 142;
Best Local Similarity 93.8%; Pred. No. 4.1e-42;
Matches 76; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MNVEHEVNLVEIHRGSKNADGKLSVKGVLFODDRCANLFEALVGTAKAKRKIVT 60
Db 62 MNVDHEVNLVEIHRGSKNADGKLSVKGVLFRDDKCANLFEALVGTAKAKRKIVT 121
Qy 61 YAGELLQGVHDDVDIVLLQD 81
Db 122 YPGELLQGVHDDVDIVLLQD 142

RESULT 2
US-09-971-791-6
; Sequence 6, Application US/09971791
; Patent No. US20020115120A1
; GENERAL INFORMATION:
; APPLICANT: Rosanna Kapeller-Libermann